IJCSIS Vol. 8 No. 5, August 2010
ISSN 1947-5500

International Journal of Computer Science & Information Security

Editorial
Message from Managing Editor

The International Journal of Computer Science and Information Security is a monthly periodical on research in general computer science and information security.

Target Audience: IT academics, university IT faculties; industry IT departments; government departments; the mobile industry and the computing industry.

Coverage includes: security infrastructures, network security: Internet security, content protection, cryptography, steganography and formal methods in information security; multimedia systems, software, information systems, intelligent systems, web services, data mining, wireless communication, networking and technologies, innovation technology and management.

Thanks for your contributions in August 2010 issue and we are grateful to the reviewers for providing valuable comments.

Available at http://sites.google.com/site/ijcsis/

IJCSIS Vol. 8, No. 5, August 2010 Edition

ISSN 1947-5500 © IJCSIS 2010, USA.

Abstracts Indexed by (among others):

IJCSIS EDITORIAL BOARD

TABLE OF CONTENTS

Blind Multiuser Detection in SDMA-aided MIMO OFDM Systems by FastICA Algorithm

Yu Du, Kang Yen, Yechang Fang and Nansong Wu
Department of Electrical and Computer Engineering
Florida International University
Miami, Florida, United States
E-mail: {dyu002, yenk, yfang003, nwu001}@fiu.edu

***Abstract*—The techniques of MIMO (Multiple Input and Multiple Output) OFDM (Orthogonal Frequency Division Multiplexing) systems have been adopted by IEEE 802.11n wireless communication standard (Wi-Fi). Multiuser Detection by blind signal separation (BSS) techniques in MIMO OFDM systems recently attracts researchers' interests by their various advantages. The performance of a MIMO OFDM system mainly depends on reliability and effectiveness of the system, and the measurement of reliability is bit error ratio (BER). On the premise of effectiveness, a new blind multiuser detection approach based on FastICA (fast fixed-point algorithm for independent component analysis) is proposed for a SDMA-aided (Spatial Division Multiple Access) MIMO OFDM system with additive white Gaussian noise (AWGN) for the target of better reliability. Compared with the usually-used spatial-multiplexing-based MIMO OFDM system, the proposed method offers better BER performance. Simulation results confirm the better reliability of the proposed approach.**

Keywords- Blind Multiuser Detection; SDMA; MIMO OFDM; FastICA

I. INTRODUCTION

The techniques used in MIMO (Multiple Input and Multiple Output) OFDM (Orthogonal Frequency Division Multiplexing) systems have been adopted by IEEE 802.11n wireless communication standard (Wi-Fi) [1], [2]. For any receiver of MIMO OFDM systems, some priori parameters such as characteristics of users and noise, timing information of users and noise, relative amplitude and training sequences are needed [3]. Blind multiuser detection is a novel approach without retransmitting them in MIMO OFDM systems.

Blind signal separation algorithms adopted in MIMO OFDM systems usually contain two classifications, which are second order statistics (SOS) and higher order statistics (HOS). With regard to SOS methods, a subspace approach with short averaging periods has been proposed [4]. Although it achieves a good estimation accuracy, the scheme requires additional information of channel orders for optimizing the ambiguity matrix, and the computation complexity is not lower when the number of subcarriers is larger than 64 [5]. In contrast with SOS, the use of HOS blind signal separation algorithms, which are also called independent component analysis (ICA) [6], decreases the influence caused by the Gaussian noise of transmission channels in MIMO OFDM systems.

The most familiar approach employing HOS in MIMO OFDM systems is Constant Modulus Algorithm [7]. Although this method provides a good solution to inter-symbol-interference (ISI) and inter-user-interference (IUI), it requires more receivers than transmitters. Another approach [8] makes use of tensor based space-time multiplexing and offers some transmission flexibility, but the problem associated with this method is that the transmitters are too complex to be implemented in practical applications. Well-established HOS methods in MIMO OFDM systems include JADE (Joint Approximate Diagonalization of Eigen-matrices) [5], FastICA [9] and Nature Gradient [10]. The JADE-LSTE (Layered Space-time Equalization) approach in [5] applies JADE to all subcarriers and uses a LSTE approach for solving the permutation and scaling indeterminacy problems. However, the limitation of this approach is its complex structure. In [9], FastICA in a MIMO OFDM system is proposed by fractional sampling in the frequency domain, and the permutation and scaling indeterminacy are solved by a PLL (phase locked loop). The fractional sampling is able to increase the number of received signals and to get better diversity at the receiver. However, this method increases the computation complexity. The Nature Gradient approach [10] is applied to each current subcarrier and the previous subcarrier has to be used sequentially to resolve the permutation and scaling indeterminacy. However, this sequential approach brings some error propagation. In general, they select similar linear schemes, and because of the higher computation and complexity of nonlinear blind signal separation algorithms, nonlinear schemes have not been included.

To realize a MIMO OFDM system, several technologies, which include beamforming, space-time coding [8], SDMA [1], spatial multiplexing [5], [7], [9] and network, have been used. Compared with other approaches, SDMA schemes are able to maximize the number of users in wireless communication systems, and the interference can be drastically reduced by users' channel impulse responses [1]. Furthermore, SDMA schemes can let multiple users in the same time or frequency or code domain to simultaneously share the same bandwidth at different locations by spatial signature, thus the capacity of the system is increased.

FastICA algorithm is capable of separating linearly mixed complex valued source signals with good computational efficiency [11]. The use of FastICA algorithm can also supply

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

enough separating accuracy for mixed complex valued source signals and has most of the advantages of other HOS blind signal separation algorithms. According to the attraction of FastICA algorithm, it is employed to the SDMA-aided MIMO OFDM system.

With enough effectiveness of a MIMO OFDM system, a smaller BER means a better reliability of the system. The objective of this paper is to apply FastICA [6], [11], to SDMA-aided MIMO OFDM systems with additive white Gaussian noise (AWGN) for detecting multiuser signals with better BER performance. The simulation results show that the proposed approach offers less BER and than the spatial-multiplexing-based MIMO OFDM systems.

This paper is organized as follows. The system model is defined in Section II. Section III describes the proposed FastICA receiver. The simulation of proposed method is presented in Section IV, followed by some conclusions in Section V.

II. System Model

A. SDMA-aided MIMO Channel

Compared with a single input single output (SISO) system which is equipped with 1 transmitting antenna and 1 receiver antenna, Figure 1 shows a generic uplink SDMA-aided MIMO system model, where each of l simultaneous users is equipped with a single transmitting antennas, while the base station's receiver exploits p-element antenna array. The scenario can improve spectral efficiency of the system, diversity gain and capacity of the system. Figure 2 [1] illustrates the schematic of an uplink SDMA-aided MIMO OFDM system. In the k^{th} subcarrier of the n^{th} OFDM symbol received by the p-element antenna array, the set of complex signal vector $\mathbf{x}[n, k]$ is constituted by the superposition of the independent signals from l mobile users and corrupted by AWGN. So the received complex signal can be expressed as [1]

$$\mathbf{x} = \mathbf{Hs} + \mathbf{n} = \bar{\mathbf{x}} + \mathbf{n} \quad (1)$$

Here the indices $[n, k]$ have been omitted for the sake of notational convenience and the vectors **x**, **s,** and **n** are given by

$$\mathbf{x} = [x_1, x_2, \cdots, x_p]^T \quad (2)$$

$$\mathbf{s} = [s_1, s_2, \cdots, s_l]^T \quad (3)$$

$$\mathbf{n} = [n_1, n_2, \cdots, n_p]^T \quad (4)$$

where T denotes transpose. The frequency domain channel transfer matrix **H** with $p*l$ dimension is given by [1]

$$\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_l]^T \quad (5)$$

where $\mathbf{H}_i$ (i = 1, 2, ⋯, l) is the set of channel transfer function vectors of the l users to each element of the p-element receiver, which is express as [1]

$$\mathbf{H}_i = [H_{i1}, H_{i2}, \ldots, H_{ip}]^T, \; i = 1, 2, \cdots, l. \quad (6)$$

There assumptions made in this model are both the complex source signal s_i and AWGN signal n_p that have zero mean and a variance of δ_l^2, and the frequency channel transfer functions H_{ij} (i = 1, 2, ⋯, l and j =1, 2, ⋯, p) are independent and stationary. The complex Gaussian distributed processes of the frequency channel transfer functions are set to zero-mean and unit variance [1].

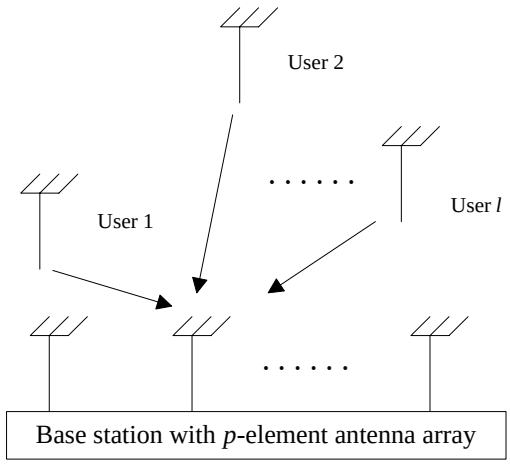


Figure 1. Generic uplink SDMA-aided MIMO system model

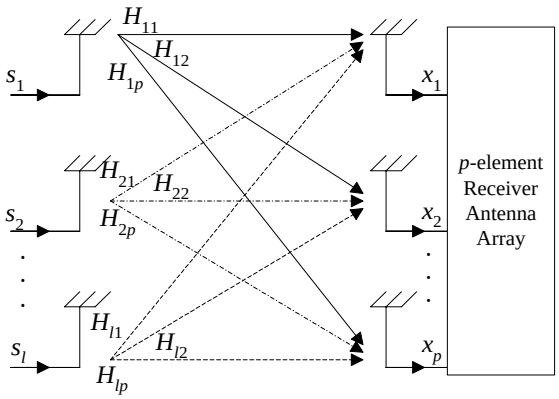


Figure 2. The schematic of uplink SDMA-aided MIMO OFDM system

B. DFT-based OFDM systems

OFDM, which is a special multicarrier modulation in the standard parallel data transmission systems, divides the frequency domain into orthogonal overlapping subcarriers. The frequency spectrums of the subcarriers can guarantee that the subcarrier signals do not interfere with each other. OFDM system alters a frequency selective broadband channel into flat fading narrowband channels so that the system maintains a high spectral efficiency.

Figure 3 shows a DFT-based OFDM transmission system and DFT technique can reduce the implementation complexity of OFDM. A sequence of discrete-time BPSK (Binary Phase

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

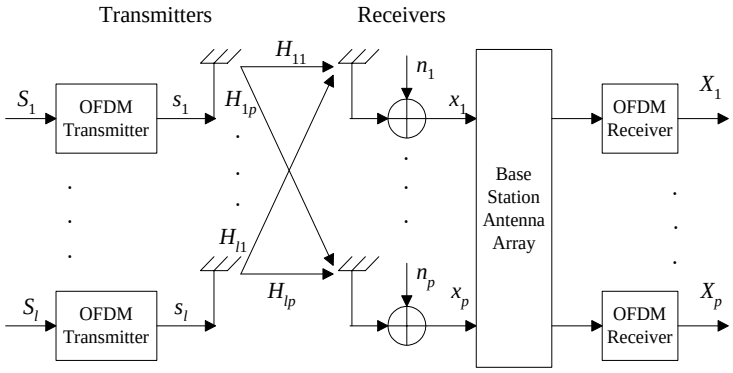


Figure 4. The SDMA-aided MIMO OFDM System Model

Shift Keying) modulated symbols goes through a serial/parallel converter, and then each block of symbols, $\mathbf{S} = [S^{(0)}, S^{(1)}, S^{(2)}, \cdots, S^{(N-1)}]^T$, is modulated by IDFT. The new vector, $\mathbf{s} = [s^{(0)}, s^{(1)}, s^{(3)} \cdots, s^{(N-1)}]^T$ is demultiplexed into a serial form to generate a single OFDM symbol. The cyclic prefix (CP) can avoid ISI, and it is inserted for each OFDM symbol at the transmitter and then is taken away at the receiver. Through the DFT demodulation at the receiver, the received frequency symbol can be shown in the following equation.

$$X^{(k)} = H^{(k)}S^{(k)} \tag{7}$$

here the factor $H^{(k)}$ is a complex valued scale at k^{th} frequency.

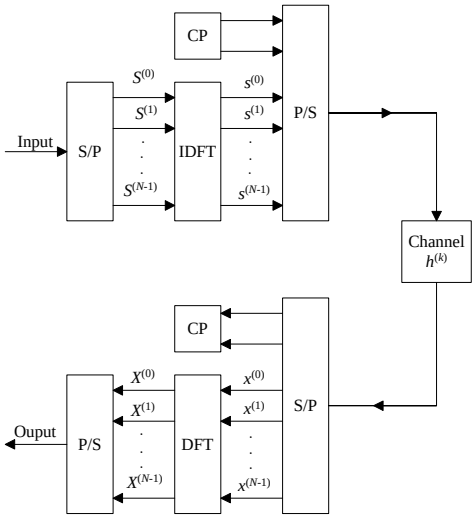


Figure 3. Block Diagram of DFT-based OFDM systems

Although OFDM scheme increases the implementation complex as shown in Figure 3 compared with a traditional serial modem, the signal in low speed subchannels is less sensitive to Inter Symbol Interference (ISI). It is worth noting that the subchannel modulation scheme is not limited to BPSK, and Quadrature Phase Shift Keying (QBSK), 16 Quatrature Amplitude Modulation (16QAM), 64 QAM and related modulation schemes have also been selected in recent years.

C. The SDMA-aided MIMO-OFDM System Model

The combination of SDMA, MIMO and OFDM may enhance the quality of wireless communication and increase the throughput of the system. The SDMA-aided MIMO OFDM is illustrated in Figure 4 [1]. The number of antennas decides the complexity of estimation at the receivers and a large number of antennas mean a complex multiuser detection. The linear SDMA-aided mixing model at k^{th} frequency can be described by (8) or in a compact expression shown in (9).

$$\begin{bmatrix} X_1^{(k)} \\ \vdots \\ X_p^{(k)} \end{bmatrix} = \begin{bmatrix} H_{11}^{(k)} & \cdots & H_{l1}^{(k)} \\ \vdots & \ddots & \vdots \\ H_{1p}^{(k)} & \cdots & H_{lp}^{(k)} \end{bmatrix} \begin{bmatrix} S_1^{(k)} \\ \vdots \\ S_l^{(k)} \end{bmatrix} + \begin{bmatrix} n_1^{(k)} \\ \vdots \\ n_p^{(k)} \end{bmatrix} \tag{8}$$

$$\mathbf{X}^{(k)} = \mathbf{H}^{(k)}\mathbf{S}^{(k)} + \mathbf{n}^{(k)} \tag{9}$$

III. THE PROPOSED FASTICA RECEIVER

In [5], it is noticeable that the ICA model for MIMO OFDM systems is an instantaneous linear mixture of each subcarrier. It is a basic requirement that allows the application of ICA to blind multiuser detection in MIMO OFDM systems. In contrast with SOS, HOS blind signal separations have better BER performance, thus the existing HOS algorithm will be employed in the proposed system. Among existing HOS algorithms, Complex FastICA algorithm [6], [11] is adopted for the proposed SDMA-aided MIMO OFDM system because of its fast convergence, less computation complexity, and high-quality BER performance.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

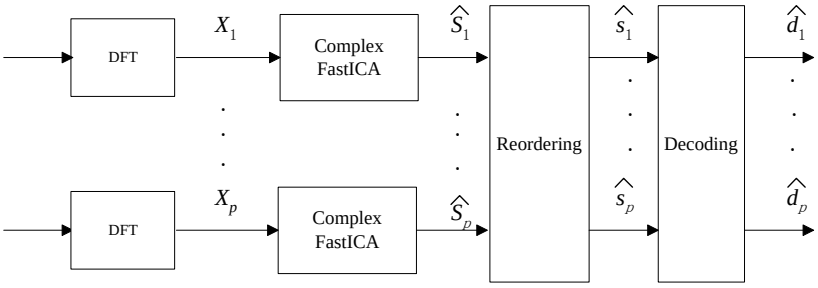


Figure 5. FastICA receiver structure

After the DFT modules of the OFDM receivers, FastICA algorithm [5] is applied. It is worth to notice that the received signals need to be whitened before using FastICA to estimate the source signals. According to [11], the minimized or maximized contrast function is

$$J_G(\mathbf{w}) = E\{G(|\mathbf{w}^H\mathbf{x}|^2)\} \tag{10}$$

where G is a smooth even function, E denotes the expectation, $\mathbf{w}$ is an n-dimensional complex weight vector, $(\cdot)^H$ means the Hermitian transpose, and $E\{|\mathbf{w}^H\mathbf{x}|^2\}=1$.

According to different performance aims, there are mainly three choices of G functions. For the better BER performance target and according to different properties of different G functions [11], we choose:

$$G(y) \approx \log(a_1 + y), \quad a_1 \approx 0.1 \tag{11}$$

where y is a complex random variable. The one-unite update matrix $\mathbf{w}$ [11] is

$$\mathbf{w}^+ = E\{\mathbf{x}(\mathbf{w}^H\mathbf{x})^* g(|\mathbf{w}^H\mathbf{x}|^2)\} - E\{g(|\mathbf{w}^H\mathbf{x}|^2) + |\mathbf{w}^H\mathbf{x}|^2 g'(|\mathbf{w}^H\mathbf{x}|^2)\}\mathbf{w} \tag{12}$$

$$\mathbf{w}_{new} = \frac{\mathbf{w}^+}{\| \mathbf{w}^+ \|} \tag{13}$$

where $(\cdot)^+$ denotes the pseudo-inverse, $(\cdot)^*$ means complex conjugation, $g(\cdot)$ is the derivative of $G(\cdot)$, g' $(\cdot)$ is the derivative of $g(\cdot)$, and $\|\cdot\|$ means the Euclidean norm. The $\mathbf{w}_{new}$ shows the update procedure. The one-unit algorithm can be extended to the whole separation matrix $\mathbf{W}^{(k)} = (\mathbf{w}_1^{(k)}, \mathbf{w}_2^{(k)}, \cdots, \mathbf{w}_p^{(k)})$ at the frequency k, and the estimation of the source signals can be found by $\mathbf{s} = \mathbf{W}^H\mathbf{x}$ at different frequencies.

Due to the permutation and scaling ambiguity of BSS methods, the reorder step has to be applied after employing FastICA algorithm. To realize this step, the pre-coding about correlation between data in different subcarriers and data in reference subcarriers is required at the transmitter, and every subcarrier can be the reference subcarrier. The pre-coding step is similar to that given in [5]

$$\mathrm{s}(k+nN) = \frac{1}{\sqrt{1+|a|^2}}[d_i(k+nN) + a d_i(k_r+nN)] \tag{14}$$

where k is the index of subcarriers, N is the total number of subcarriers, n is the index of OFDM symbols, the index $d_i(\cdot)$ is the unit variance source stream i, the constant a stands for an imaginary valued scalar, and k_r is the reference subcarrier.

The reordering module at the receiver calculates correlation between data in each subcarrier and data in the reference subcarrier. In [5], it shows that different correlations mean data in different streams and an estimation of $\mathbf{s}(n, k)$ can be obtained by the remaining subcarriers through the following equation

$$\mathbf{s}(n, k) = \mathbf{D}^{-1}(k)[\hat{s}_1(k+nN), \ldots, \hat{s}_p(k+nN)]^T \tag{15}$$

where $\mathbf{D}(k)$ is the non-singular diagonal phase shifting matrix.

The last step is to decode the original source signals. The decoded source stream i is given by [5] and has the form

$$\hat{d}_i(k+nN) = \sqrt{1+|a|^2}\,\hat{s}_i(k) - a\,\hat{s}_i(k_r+nN) \tag{16}$$

IV. SIMULATIONS

Regarding the purpose of the research and taking into account both 802.11n [2] 802.11a standard [12], the proposed system, which includes 4 transmitting antennas and a base station with 4-element antenna array, is operated in the 5GHz band, and the system bandwidth is set to 20MHz. Here the set value of transmitting antennas or antenna array is 4 instead of a large number, because increasing antenna elements can make antenna systems grow in size. Furthermore, the number of antenna elements increasing means approach a limit capacity, and it is still an open question about the maximum capacity of a limited-size antenna [13]. The total number of subcarriers is 64 and the length of cyclic prefix is 16. The bit rate of data with BPSK modulation is 16Mbit/s and OFDM symbol duration is 3.2 μs. The simulation results are average values of collected simulation data.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The BER vs. SNR performance of the proposed SDMA-aided MIMO OFDM system and conventional spatial-multiplexing-based MIMO OFDM system is shown in Figure 6. There is about 21% relative BER improvement in average when SNR range is from 5 dB to 30 dB according to Table I's data. Thus, it demonstrates that the proposed system has less BER at different SNR value and is mainly more robust than the spatial-multiplexing-based MIMO OFDM system.

TABLE I. RELATIVE BER IMPROVEMENT

SNR(dB)	5	10	15	20	25	30
Relative BER Improvement	23.1%	11.6%	20.5%	23.5%	18.8%	28.4%

V. CONCLUSIONS

In this paper, SDMA technique is applied to the MIMO OFDM system based on the analysis and discussion of different approaches and SDMA is speculated to be one of the most capable ways in improving performance of MIMO OFDM systems. A significant measurement of reliability for MIMO OFDM system, bit error rate, is considered. For better bit error rate performance, FastICA algorithms are demonstrated and one of them is selected to such a target.

A possible simulation structure is designed for proposed system. And the BER performance of the proposed scenario is analyzed and compared with spatial-multiplexing-based MIMO OFDM system. The simulation results show that the proposed method has better BER performance than spatial-multiplexing-based MIMO OFDM system. Furthermore, the simulation results also illustrate that the complex FastICA algorithm indicates its validity for SDMA-aid MIMO OFDM systems.

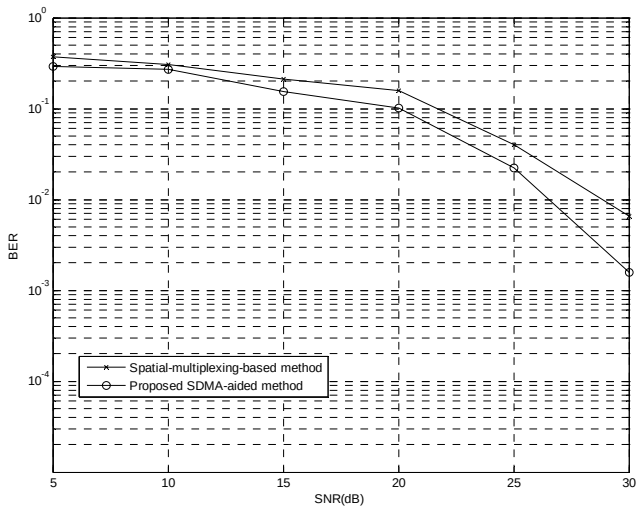


Figure 6. BER performance of the proposed system and spatial-multiplexing-based method

REFERENCES

[1] M. Jiang and L. Hanzo, "Multiuser MIMO-OFDM for Next-generation Wireless Systems," Proceedings of the IEEE, Vol.95, No.7, pp.1430-1469, July 2007.

[2] Wi-Fi CERTIFIED 802.11n, "Draft 2.0: Longer-Range, Faster-Throughput, Multimedia-Grade Wi-Fi Networks," IEEE 802.11n Draft, June 2007.

[3] J. Ma, Y. Niu and H. Chen, "Blind signal separation," National Defense Industry Press, China, 2006.

[4] C. Tu and B. Champagne, "Subspace Blind MIMO-OFDM Channel Estimation with Short Averaging Periods: Performance Analysis," 2008 IEEE Wireless Communications and Networking Conference, pp.24-29, March 31 - April 3, 2008.

[5] L. Sarperi, X. Zhu, A. K. Nandi, "Blind OFDM receiver based on independent component analysis for Multiple-Input Multiple-Output systems," IEEE Transactions on Wireless Communications, Vol.6, pp.4079-4089, November 2007.

[6] A. Hyvarinen, J. Karhunen and E. Oja, "Independent Component Analysis," Wiley-Interscience, 2001.

[7] T. Abrudan, M. Sirbu and V. Koivunen, "Blind Multi-user Receiver for MIMO-OFDM Systems," the 4th IEEE Workshop on Signal Processing Advances in Wireless Communications, pp.363-367, June 2003.

[8] A. Almeida, G. Favier, C. Cavalcante and J. Mota, "Tensor-Based Space-Time Multiplexing Codes for MIMO-OFDM Systems with Blind Detection," the 17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, pp.1-5, Sept. 2006.

[9] S. R. Curnew and J. Ilow, "Blind Signal Separation in MIMO OFDM Systems Using ICA and Fractional Sampling," ISSSE'07 International symposium on Signals, Systems and Electronics, pp.67-70, July 30 -Aug. 2, 2007.

[10] B. Guo, H. Lin and K. Yamashita, "Blind signal recovery in multiuser MIMO-OFDM System," the 47th IEEE International Midwest Symposium on Circuits and Systems on Vol.2, pp.637-640, July 2004.

[11] E. Bingham and A. Hyvarinen, "ICA of complex valued signals: a fast and robust deflationary algorithm," IEEE Proc. of the International Joint Conference on Neural Networks, vol.3, pp.357-362, July 2000.

[12] IEEE 802.11a, "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: High-speed Physical Layer in the 5GHz Band," supplement to IEEE 802.11 Standard, Sept. 1999.

[13] I.P. Kovalyov, "SDMA for Multipath Wireless Channels - Limiting Characteristics and Stochastic Models," Springer Verlag Berlin Heidelberg New York, Germany, 2004.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A Hierarchical Resource Allocation Architecture for Mobile Grid Environments

S. Thenmozhi
Assistant Professor, Department of Computer Applications
Chettinad College of Engineering and Technology
Tamilnadu, India.
sthenmozhiphd@gmail.com

A.Tamilarasi
Professor, Department of MCA
Kongu Engineering College, Erode
Tamilnadu, India

***Abstract*— The mobility issue in grid environments has established new challenges to the research communities particularly in the areas of scheduling, adaptation, security and mobility. Especially, the resource allocation becomes more challenging when mobility is considered in grid environment. Hence it is necessary to consider the mobility of users along with the resource availability while scheduling the resources for the execution of jobs. In this paper, we propose to design a Hierarchical Resource Allocation Architecture (HRAA) which includes resource monitoring and scheduling operations for mobile grid. In this architecture, the Mobile Grid is divided into clusters. Each cluster has one cluster head (CH). A master server (MS) controls each local clusters and has frequent updates of all the CH information. Each CH has a monitoring agent (MA) which will periodically predict the mobility of the cluster nodes and monitor the resource availability and update their values. When the MS forwards the job request of a user to the ideal CH, the CH schedules the jobs based on the predicted time for resource availability and sufficiency of the resources. By simulation results, we show that our proposed architecture achieves good success ratio and throughput with reduced delay and energy consumption.**

Keywords-Resource Allocation, monitoring agent (MA), Hierarchical Resource Allocation Architecture (HRAA), cluster head, master server (MS)

I. INTRODUCTION

A. Mobile Grid

The Grid is a distributed, high performing computing and data handling infrastructure. It provides common interfaces for all the resources by using standard, open, general-purpose protocols and interfaces by incorporating the geographically and organizationally dispersed, heterogeneous resources. But it is the basis and the enabling technology for the persistent and utility computing [1]. Multiple administrative domains, autonomy, heterogeneity, scalability and dynamicity/adaptability are the important features of the Grid.

The mobility issues are handled by enabling both fixed and mobile users, in the mobile grid environment. By using the underlying technologies transparently and efficiently, the access for both fixed and mobile grid resources are provided. Mobile Grid is derived from Grid with the additional support of mobile users and resources in a seamless, transparent, secure and efficient way. Moreover, it has the capability to organize the underlying ad hoc networks. It forms arbitrary and unpredictable topologies by providing a self-configuring grid system of mobile resources which are connected by the wireless links [1].

The mobile grid uses the advanced capabilities of wireless networks and lightweight thin devices. Though grid computing integrates geographically dispersed resources and users to create a dynamic virtual organization, most of the resources are static in nature. The user and the resource participating in problem solution are the two basic units of the processing environment. [2].

For many large scale applications which are dynamic in nature and require transparency for users, Mobile Grid is considered as the best solution. Grid will increase the job throughput and performance of the corresponding applications by applying efficient mechanisms for resource management. Moreover, it will enable the advanced forms of cooperative work by allowing the seamless integration of resources, data, services and ontologies [1]. Some of the applications of the mobile grids are scientific, public services and commercial businesses. Mobile grids integrate the mobile devices like laptops, PDAs (Personal Digital Assistants) [2].

B. Resource Allocation (or) Management in Mobile Grid

Resource allocation is a basic issue to achieve high performance on a grid workflow [3]. Resource allocation can be classified into resource selection (discovery) and resource binding (acquiring). The resource selection is separated from resource binding based on the common architecture of conventional resource brokering systems. These systems mainly concentrate on the resource selection for providing complex resource specification languages and resource selection algorithms. On providing a resource specification by using available resource information, a resource selection algorithm first discovers a matching set of resources and negotiates with an individual local resource manager. Then the application attempts to acquire the resources [4].

C. Resource Allocation challenges in Mobile Grid

When mobility is considered, the resource selection becomes more challenging. Therefore, it is necessary to consider the mobility of users with the resources in resource

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

selection. The mobility issue in grid environments has established new challenges to the research communities particularly in the areas of scheduling, adaptation, security and mobility. Mostly, the behavior of the user and/or mobile device is highly unpredictable which produces disconnection problems

In grid environments, new challenges are introduced to the research communities especially in the fields of mobility, scheduling, adaptation and security due to issues like mobility, power consumption and size of devices. The secondary problems are the small screen size and difficult input mechanisms. Peer-to-peer computing provides many useful technologies and ideas, for developing scalable and reliable mobile grids. In mobile environment, the most challenging problem is the disconnection problem [2].

The problems of resource allocation are:

1) Identification of an appropriate service and resources.
2) Based on certain criteria, allocating the resources, such as pricing or priority.
3) Dynamically allocating and updating the state of the resources [5].

A centralized allocation manager is not possible because the portions of the Grid may apply different allocation strategies due to decentralized Grid policies. The lack of accurate resource status information at the global scale is the additional challenge to the Grid resource allocation. At their removal, the knowledge of real time environment has been limited by the allocation strategies which are utilized by the users and brokers. Therefore the possible allocation mechanisms should not depend on the availability of current global knowledge [6].

In this paper, we develop a resource monitoring and scheduling scheme for mobile grid. In this scheme, the user submits the job request for a job to be executed to a server. The job request contains job description, number of resources required, expected job completion time and the quoted price allotted for it.

II. Related Work

Xiaozhi Wang et al [7], have proposed a layered structure of Grid QoS. Based on the analysis of the content of grid resource allocation management (GRAM) based on QoS, their work puts forward the architecture of GRAM based on QoS. Through mapping, converting and negotiating the QoS parameters, it can implant the user's requirement about QoS in the process of resource allocation management, and connect Grid QoS with GRAM very well. All these provide a reasonable consulting model for QoS and resource allocation management in grid. Their work raised the performance of GRA from different aspects. In their process of searching, system may negotiate with the user, then get final result: not being able to supply, being able to supply or reducing QoS demands to supply.

Konstantinos Katsaros et al [8] have discussed a campus-wide hierarchical Mobile Grid system architecture in which mobile nodes (MNs), willing to offer their computational resources, move between WLANs. This willingness of the mobile node is based on reciprocity criteria. They have also considered the divisible load applications (DLA) which divides the load of computation into parts and made to carry out independently. Moreover, they have described an architecture for the realization of a Mobile Grid and investigate key design decisions and optimizations.

Lei Zhang et al [9] used the Particle Swarm Optimization (PSO), the latest evolutionary optimization technique to solve the task scheduling problem in grid (computational grids) environment. Here each particle is represented as a possible solution, and the position vector is transformed from the continuous variable to the discrete variable. They have also aimed to generate an optimal schedule to get the minimum completion time while completing the tasks.

Abdullah M. Elewi et al [10] have addressed the problem of energy efficient real time task scheduling where the tasks are dependent due to exclusive access shared resources. Moreover, they have proposed about the enhancements over the existing dual speed switching algorithm (DSA) where their proposed algorithm achieves more energy saving and has the capability to function with both SRP and DPCP protocols.

Homam Reda El-Taj et al [11] have given a survey about mobile computing, the mobile grid computing, mobile agents and how to apply mobile agents on mobile grid computing and what has been done to solve the issues in these areas of study.

Ming Wu et al [12] have proposed a prototype of Grid Harvest Service (GHS) which provide dynamic and self-adaptive task scheduling. Their study is made upon task scheduling of a parallel or distributed application with a divisible workload in a heterogeneous environment. It also shows the possibility of integrating the three parts of task scheduling, that is the task allocator, scheduler and predictor into existing toolkits for better service.

Hesham A. Ali et al [13] have introduced a "self ranking algorithm", which will be used to build a mobile computing scheduling mechanism to schedule the tasks on the mobile devices, which will maximize the profit of the mobile devices which are integrated within the grid using their computational power as an addition to the system overall power.

Gurleen Kaur et al [14] have addressed the promising and bright side of the grid computing technology. They have explored the grid capabilities further by organizing the grid computing concept from two broad perspectives such as User's Perspective and Administrator's Perspective.

III. Hierarchical Resource Allocation Architecture

A. System Design

In our system design, a set of machines and a cluster head (CH) are included in each local cluster. A master server (MS) controls and groups many local clusters. The MS collects information about the resources in its local clusters. Then it stores the information in its own database.

The proposed scheme divides a given task that is submitted by a user into subtasks. Then it finds spare processors and other critical resources on the network, distributes the

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

subtasks, monitors the progress of the subtasks, and restarts the subtasks that fail.

It consists of the following steps:

1) The MS divides a given task into a sequence of subtasks and allocates the subtasks to the local cluster heads (CH).
2) The CH finds the available processing power and average mobility on the local machines and then distributes the subtasks over these machines.
3) The CHs gather the completed subtasks from the local machines and then send the data back to the MS.
4) The MS aggregates the completed subtasks and then stores the results in its own database, and
5) MS also sends back the results to the application of the respective users.

B. Mobility Metric

User Range is a given fact about the initiator coverage, in which user can communicate with mobile devices. Average Mobility, a derived parameter, represents the average mobility of a resource and/or user (based on user and resource mobility). Average mobility is calculated based on two recent communications between user/initiator and resource with respect to the user/initiator.

Average Mobility is calculated as

$$Average\ \ Mobility = f1 - f2 \tag{1}$$

Where, $f1$ and $f2$ are the first history and second history respectively. The history is simply the distance between user and resource and it can be calculated by finding difference between the two recent interactions.

$Ptime$ shows the predicted time for resource availability within the user's range and is calculated by the following equation

$$Ptime = (\text{User Range} - \text{Distance}) / \text{Average Mobility} \tag{2}$$

The "Distance" is the net difference between the new locations of user and resource.

C. Resource Availability Metric

The monitoring agents (MA) estimates the workload of its grid nodes $(n_i, i = 1,2,\cdots,k)$ present in the cluster CL_1 using the following formula:

$$WL_i = \frac{CWL_i + \left(\sum_{j=1}^{k} Jobsize_j\right)}{Power_i} \tag{3}$$

Where CWL_i is the current workload of n_i, WL_i is the work load of n_i

$Power_i$ is the power of the node n_i and $Jobsize_j$ is the size of the Job_j.

D. Scheduling Strategy

After estimating the $Ptime$ and WL values, the MA sends these values to its cluster head CH_1. The CH_1 then schedules the jobs if their grid node satisfies the following condition

$$If\ (Ptime / WL) > Th \tag{4}$$

where Th is a threshold value (which can be fixed based on the job request).

From 4, we can observe that if the $Ptime$ is less and if the work load is more then the gird nodes are unable to execute the job request. Therefore, the jobs are executed by the nodes only when $Ptime$ is high and the work load is low. If the CH is unable to allocate the resources in its cluster, it resubmits the job request information to the MS.

E. Functions of HRAA

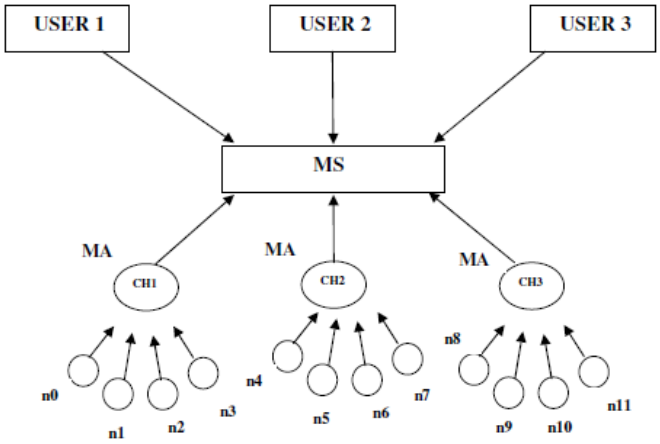


Figure 1. Functions of HRAA

The sequence of operations in HRAA is shown in Figure. 1. In this figure, the arrows represent the communication messages and the nodes represent the agents/servers. The sequence is as follows.

1) The MAs of each node in the local cluster send the resource status information to the CHs.
2) The CHs send this information to the MS.
3) The MS then create a database which contains information about the status and the price of each resource.
4) A user submits its job details and the resource requirements to the MS.
5) The MS sends the job request information to the local CH.
6) The local CH allocates the resources in their control depending on the predicted time, the average processing power and average load of its local machines

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

7) If any CH unable to allocate the resources in its cluster, it resubmits the job request information to the MS.

8) The MS forwards the job request information to other CH. The process is continued until the job is successfully assigned

9) The CHs gather the completed subtasks from the local machines and then send the data back to the MS.

10) The MS aggregates the completed subtasks and then stores the results in it's own database.

11) MS also sends back the results to the application of the respective users.

IV. Simulation Results

A. Simulation Model and Parameters

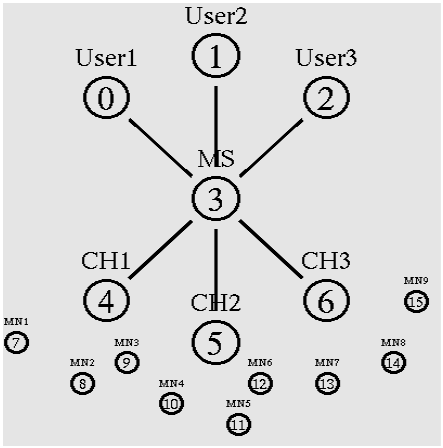


Figure 2. Simulation Setup

In this section, we examine the performance of our Hierarchical Resource Allocation Architecture (HRAA) with an extensive simulation study based upon the ns-2 network simulator [15]. The simulation topology is given in Figure 2. We compare our results with Agent-based Resource Allocation Model (ARAM) [5].Various simulation parameters are given in table 1.

Table I. Simulation Settings

Mobile Nodes	9
Users	3
Clusters	3
Area Size	1000 X 1000
Mac	802.11
Radio Range	250m
Routing Protocol	DSDV
Simulation Time	50 sec
Traffic Source	CBR
Packet Size	512
Rate	250kb,500kb,….1000kb
No. Of tasks	2,4,6,8 and 10
Speed	5m/s
Transmit Power	0.660 w
Receiving Power	0.395 w
Idle Power	0.335 w
Initial Energy	10.1 J

B. Performance Metrics

In our experiments, we measure the following metrics.

Average Execution Delay: It measures the average delay occurred while executing a given task.

Average Success Ratio: It is the ratio of the number of tasks executed successfully and the total number of tasks submitted.

Average Energy: It is the average energy consumption of all nodes in executing the tasks.

Throughput: It is the number of tasks finished successfully.

C. Results

A. Based on Rate

In this experiment, we vary the execution rate as 250Kb, 500Kb, 750Kb and 1000Kb.

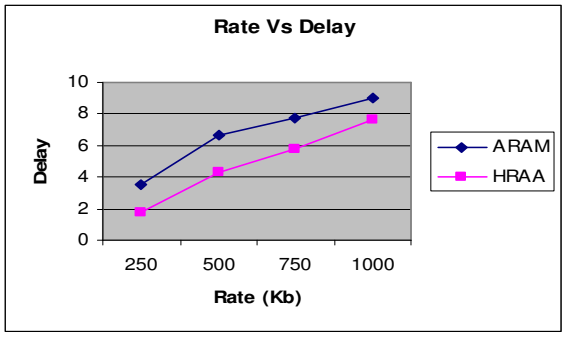


Figure 3. Rate Vs Delay

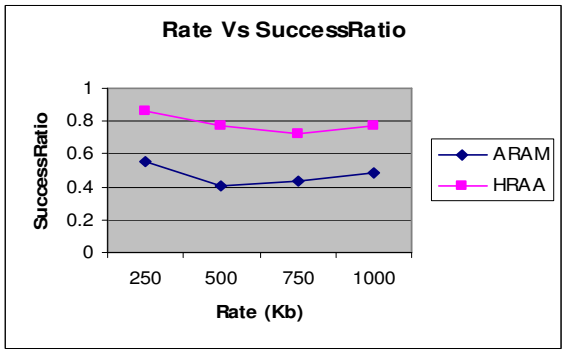


Figure 4.Rate Vs Success Ratio

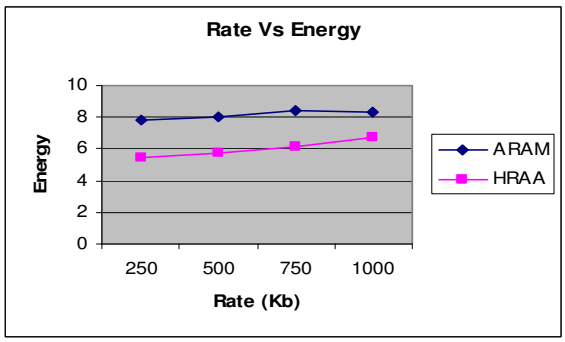


Figure 5.Rate Vs Energy

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

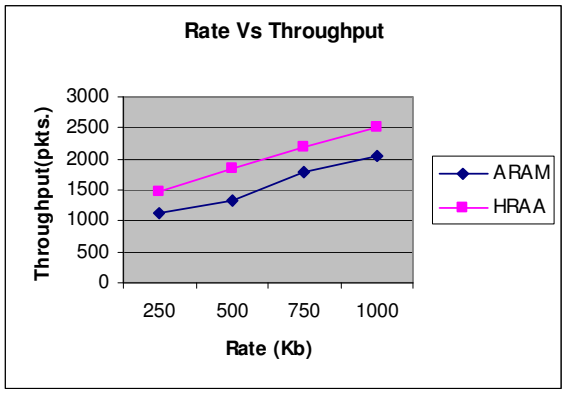


Figure 6.Rate Vs Throughput

From Figure 3, it is clear that when the execution rate is increased then the delay also increases. We can see that the average execution delay of the proposed HRAA algorithm is less when compared to the ARAM algorithm when the rate is increased.

Figure 4 shows that when the execution rate is increased then the success ratio gets decreased. From the figure we can see that the HRAA achieves good success ratio, compared to ARAM.

Figure 5 shows that when the execution rate is increased then the energy consumption gets increased slightly. From the results, we can see that HRAA consumes less energy than the ARAM.

Figure 6 shows that when the execution rate is increased then the throughput is also increased. As we can see from the figure, the throughput is more in the case of HRAA when compared to ARAM.

B. Based on Number of Tasks

In this experiment, we vary the number of tasks to be executed as 2, 4….10.

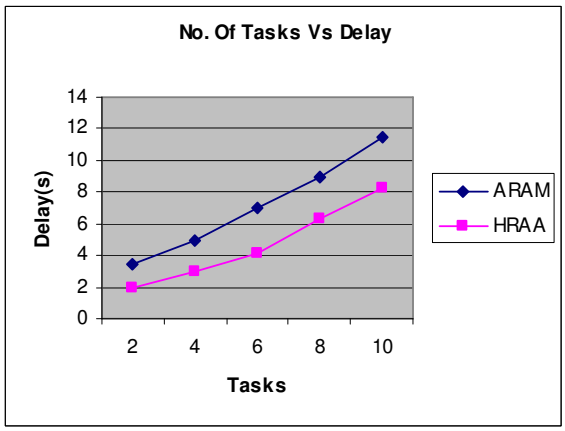


Figure 7. Number of Tasks Vs Delay

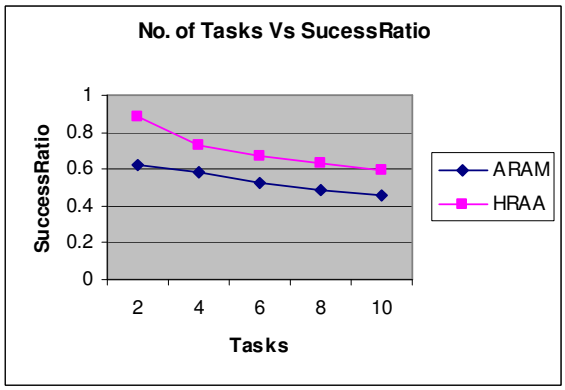


Figure 8. Number of Tasks Vs Success Ratio

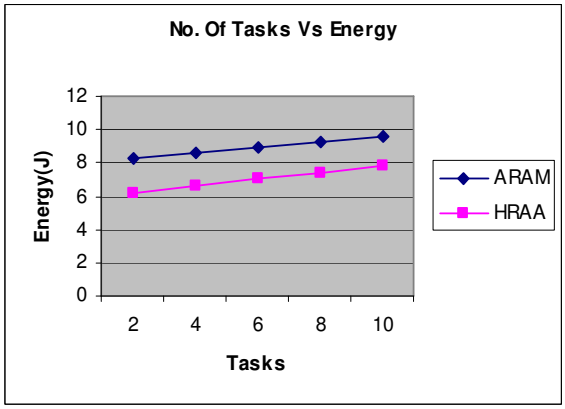


Figure 9. Number of Tasks Vs Energy

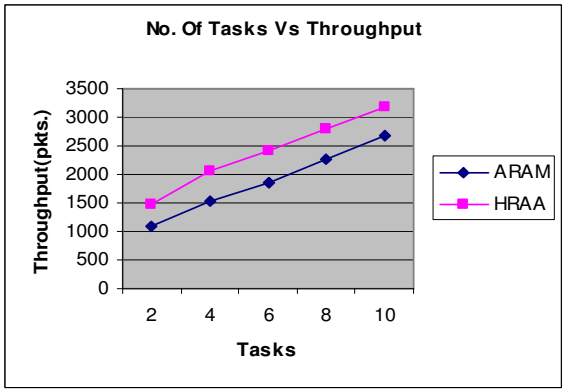


Figure 10. Number of Tasks Vs Throughput

From Figure 7, it is clear that when the number of tasks is increased then the delay also gets increased. We can see that the average execution delay of the proposed HRAA algorithm is less when compared to the ARAM algorithm when the number of tasks is increased.

Figure 8 shows that when the number of tasks is increased then the success ratio gets decreased. From the figure we can see that the HRAA achieves good success ratio, compared to ARAM.

Figure 9 shows that when the number of tasks is increased then the energy consumption gets increased slightly. From the

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

results, we can see that HRAA consumes less energy than the ARAM.

Figure 10 shows that when the number of tasks is increased then the throughput is also increased. As we can see from the figure, the throughput is more in the case of HRAA when compared to ARAM.

V. CONCLUSION

In this paper, we have designed a Hierarchical Resource Allocation Architecture (HRAA) which includes monitoring and scheduling operations for mobile grid. In this architecture, a set of machines and a cluster head (CH) are included in each local cluster. A master server (MS) controls and groups many local clusters. The MS collects the information about the resources in its local clusters and stores it in its own database. Each CH has a monitoring agent (MA) which will periodically predict the mobility of the cluster nodes and monitor the resource availability and update their values. The MS forwards the job request of a user to the ideal CH. If the CH finds the available processing power, workload and predicted time for resource availability on the local machines, then it distributes the subtasks over these machines. Otherwise the job request is resubmitted to the MS which again forwards the job request to another CH. The process is continued until the job is successfully assigned. The completed job is returned back to the MS through the corresponding CH. The MS then returns it to the requested user. By simulation results, we have shown that our proposed protocol achieves good success ratio and throughput with the reduced delay and energy consumption.

REFERENCES

[1] Litke, A., Skoutas D. and Varvarigou T., "Mobile Grid Computing: Changes and Challenges of Resource Management in a Mobile Grid Environment", presented in Workshop: "Access to Knowledge through Grid in a Mobile World", PAKM 2004 Conference, Vienna

[2] Umar Farooq, Saeed Mahfooz and Wajeeha Khalil, "An Efficient Resource Prediction Model for Mobile Grid Environments", 2006- PG Net.

[3] Pengcheng Xiong, Yushun Fan, "Cost-aware Grid Workflow Resource Allocation", IEEE Computer Society, 2007.

[4] Yang-Suk Kee, Ken Yocum, Andrew A. Chien, Henri Casanova, "Improving Grid Resource Allocation via Integrated Selection and Binding", 2006 IEEE

[5] S.S. Manvia, M.N. Birjeb, Bhanu Prasad, "An Agent-based Resource Allocation Model for computational grids", IOS Press Amsterdam, The Netherlands, 2005.

[6] Aram Galstyan, Karl Czajkowski,Kristina Lerman, "Resource Allocation in the Grid Using Reinforcement Learning", IEEE Computer Society 2004.

[7] Xiaozhi Wang, Junzhou Luo, "Architecture of Grid Resource Allocation Management Based on QoS", Springerlink – 2004.

[8] Konstantinos Katsaros and George C. Polyzos , "Optimizing Operation Of A Hierarchical Campus-Wide Mobile Grid", The 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'07)

[9] Lei Zhang, Yuehui Chen, Runyuan Sun, Shan Jing and Bo Yang, "A Task Scheduling Algorithm Based on PSO for Grid Computing ", International Journal of Computational Intelligence Research. (2008),

[10] Abdullah M. Elewi, Medhat H. A. Awadalla and Mohamed I. Eladawy , "Energy Efficient Real Time Scheduling Of Dependent Tasks Sharing Resources" Proceedings of the High Performance Computing & Simulation Conference, 2008.

[11] Homam Reda El-Taj and Chan Huah Yong, "Applying Mobile Agents on Mobile Grid Computing", Computer Science Postgraduate Colloquium (CSPC'07).

[12] Ming Wu and Xian-He Sun, "A General Self-adaptive Task Scheduling System for Non-dedicated Heterogeneous Computing", Proceedings of IEEE International Conference on Cluster Computing, 2003.

[13] Hesham A. Ali and Tamer Ahmed Farrag, "High Performance Mobile Computing Algorithm Based On Self-Ranking Algorithm (Sar)", Department of Computers and Systems, Faculty of Engineering, Mansoura University, Egypt, Department of Computers and Systems, Faculty of Engineering, Mansoura University, Egypt

[14] Gurleen Kaur and Inderpreet Chopra, "Grid Computing – Challenges Confronted and Opportunities Offered", Proceedings of Challenges and Opportunities in Information Technology, COIT-2007.

[15] Network Simulator: http:// www.isi.edu/nsnam/ns

S. Thenmozhi received MCA degree from Annamalai University, India in 2000 and M.Phil degree in computer science from Bhrathiar University, India in 2003. Currently she is pursuing the Ph.D from Anna University Coimbatore, India. Her area of interest in research includes Grid Computing and Mobile Computing. She has published 6 papers in National/International Conferences/Journals. Presently, she is working as Assistant Professor in Department of Computer Applications, Chettinad College of Engineering and Technology, Tamilnadu, India.

A.Tamilarasi post graduated from Bharathiar University, India 1986. She obtained her Ph.D from University of Madras, Chennai in the year 1994. She was awarded JRF by UGC in the year 1986. She has published more than 40 research papers in the reputed national/international Journals. She is author of 10 books. Her area of interest includes Semigroup theory, Soft computing, Data Mining. Presently she is working as a Professor Department of MCA, Kongu Engineering College, Erode, Tamilnadu, India

An Efficient Detection and Management of False Accusations in Ad Hoc Network

Yunho Lee , Soojin Lee

Department of Computer & Information Science,
Korea National Defense University,
Seoul, South Korea
yunholee@gmail.com, cyberkma@gmail.com

Abstract

Since ad hoc networks rely on the cooperation of all the participating nodes for routing and forwarding, the fast detection of malicious nodes is a critical issue. Therefore, the dissemination of observed behavior information of neighboring nodes is efficiently used for detecting misbehaving nodes. However, this may make ad hoc networks vulnerable to false accusation.

In this paper, to detect quickly and manage the false accusations of malicious nodes in the hierarchical ad hoc network such as military tactical ad hoc network, we propose a new efficient way using a Node Weight Management Server (NWMS). The NWMS which is the upper layer node maintains a weight value for every node in their area and detects and isolates malicious nodes using the weight value of nodes. In addition, our system provides a rescuing method for incorrectly imposed weight values. By means of simulation we have evaluated the efficiency of our approach for detecting and managing misbehaving nodes. The simulation results indicate that proposed mechanism is significantly efficient for handling misbehaving nodes.

Keywords; Ad hoc, false accusation, NWMS

Ⅰ. Introduction

Since ad hoc networks have no fixed infrastructure and can be deployed fast, they can be applied to various fields such as military tactical operations, emergency situation, rescue mission and establishment of temporal conference. Many technical research related with this field have been proposed. Early research effort assuming a friendly relationship and cooperation between nodes mainly focused on developing more efficient routing protocol. In recent years, security has become a primary concern to provide security services, such as confidentiality, integrity, authentication, and availability, to ad hoc nodes or users [1, 2, 3, 4, 5].

Although security has long been an active research issue in ad hoc network, many new challenges and opportunities have been posed by the unique characteristics of the ad hoc network such as open peer-to-peer network architecture, resource limitations, shared wireless medium, and highly dynamic network topology. Moreover, since the existing ad hoc routing and MAC protocols assume a trust relationship and cooperation between mobile nodes, misbehaving nodes may cause the performance degradation of the network as well as the energy consumption of normal nodes. In the worst case, the network can be partitioned. Therefore, enhancing the security is critical issue in ad hoc network.

The two most commonly used approaches to enhance the security in ad hoc networks are prevention, the detection and reaction mechanisms. Prevention mechanisms [6, 7, 8, 9, 10, 11] pursue the object that only friendly and mutually trusted nodes are included into the routing path by using the cryptography algorithm, key management mechanism, and one-way hashing chains. Although ad hoc networks are initially constructed by trusted nodes using the prevention mechanisms, some node could be compromised by adversaries that may use counterfeit information to breakdown the network and conserve their own resources. Prevention mechanisms, by themselves cannot ensure complete cooperation among nodes in the network. Most of vulnerabilities and the attacks in the ad hoc have been the result of bypassing prevention mechanisms.

Therefore, detection and reaction mechanisms [12, 13] are essential in ad hoc networks. Most existing studies associated with this research are based upon the detection technique of particular selfish nodes which do not provide packet forwarding to conserve their resources. But there is little or no research to resolve the problem of bogus information produced by malicious nodes which intentionally identify a normal node as a malicious node. Therefore, we focus our intention on the false accusation problem of malicious nodes in the detection and reaction mechanisms.

In this paper, we consider tactical ad hoc networks [14, 15] as a hierarchical architecture and then set the upper layer node as a *Node Weight Management Server (NWMS)*. Main tasks of the NWMS are the weight maintenance of suspected nodes which are detected and reported by neighboring nodes, and the decision to the isolation of suspected node or nodes.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The major contributions of our paper are summarized as follows:

a) Our proposed mechanism first takes care of the problem concerning the false accusation of malicious nodes in the tactical ad hoc network.

b) Our proposed mechanism can keep track of misbehaving nodes by using Misbehaving Node List (MNL) maintained by the NWMS.

c) Our proposed mechanism also has the function of relieving the misidentification of normal nodes that may be caused by a temporary network error. Although normal nodes were pointed out as misbehaving nodes, those nodes can be relieved as reducing the weight value in a Suspect Node List (SNL) maintained by each node.

The rest of the paper is organized as follows. Related work is discussed in Section Ⅱ. We present assumptions and some background knowledge in Section Ⅲ. We present our proposed mechanism which can handle the false accusation of malicious nodes efficiently by using the NWMS in Section Ⅳ. Section Ⅴ shows the performance evaluation in simulation results. Finally, we conclude the paper in Section Ⅵ.

Ⅱ. Related Work

Ad hoc network works properly only if the participating nodes cooperate in routing and forwarding. However, due to the resource-limitation of each node, it may be advantageous for individual node not to cooperate. This causes a serious problem in the wireless network because each node needs to cooperate between each other. Some node takes selfish actions such as only receiving other node's service but not utilizing their resources to prolong their longevity or to achieve malicious aims, such as performance degradation and network partition, malicious nodes may make false accusations which intentionally point out a normal node as a malicious node.

In this section we give a brief introduction about two mechanisms to enhance the security proposed for the ad hoc network; prevention, detection and reaction mechanisms. And then, we discuss the limitations of these mechanisms.

Prevention mechanisms; Stajano and Anderson [6] authenticate users by 'imprinting' in analogy to ducklings acknowledging the first moving subject they see as their mother. Imprinting is realized by accepting a symmetric encrypting key from the first device that sends such a key. However, a drawback of this paper is that it does not address routing or forwarding problems that may occur. The Secure Routing Protocol by Papadimitratos and Hass [7] guarantees correct route discovery, so that fabricated, compromised, or replayed route replies are rejected or never reach the route requester. However, this protocol has basically a handicap that assumes a security association between end-points of a path. *Ariadne*, a secure on-demand routing protocol by Hu, Perring, and Johnson [8], prevents attackers from tampering with uncompromised routes consisting of friendly nodes. It is based on Dynamic Source Routing [9] and relies on symmetric cryptography only. It uses a key management protocol called TESLA that relies on synchronized clocks, which is, arguably, an unrealistic requirement for ad hoc networks. Sanzgiri, Dahill, Levine, Belding-Royer proposed *ARAN*[10], a routing protocol for ad hoc networks that uses authentication and requires the use of a trusted certificate server. However, this mechanism is vulnerable to reply attacks using error messages unless the nodes have time synchronization. *Secure DSR* by Kargl, Geiss [11], is recently secure routing algorithms that counter various attacks such as forging, modifying, or dropping of routing message. However, this mechanism only protects the control plane, but do not secure the forwarding of data messages.

As aforementioned, these prevention mechanisms for ad hoc networks have only focused on providing secure routing functionality. In addition, these mechanisms based on correlation between participated nodes, some compromised node may cause critical problems such as network partition or breakdown. Therefore, detection and reaction mechanisms are essential in ad hoc networks.

Detection and reaction mechanisms; Marti, Giuli, Lai and Baker [12] propose *watchdog* and *pathrater* mechanism to mitigate routing misbehavior. This mechanism is employed by each node individually to observe the message sent by neighboring nodes. Watchdog mechanism relies on overhearing the communication of neighboring nodes. If *watchdog* identifies misbehaving nodes, *pathrater* helps routing protocols avoid these nodes. However, this mechanism has limitations. First, the detected misbehaving node is not punished. In other words, the nodes rely on their own watchdog exclusively and do not exchange the observed behavior information of neighboring nodes with others. Second, *pathrater's* function, search the bypass route avoid the misbehaving nodes, which allows these rogue nodes to conserve energy. Third, whenever the misbehaving nodes want to send their message, they can join the network thus making it attractive to deny cooperation. Buchegger, Boudec [13] propose a protocol, called CONFIDANT, for making misbehavior unattractive. CONFIDANT consists of the several components; *Neighborhood Watch* for observations, *Trust Manager* to deal with incoming and outgoing alarm messages, *Reputation System* to record reputations about first-hand and trusted second-hand information *and Path Manager* for path re-ranking and deletion of paths containing malicious nodes. It aims at detecting and isolating misbehaving nodes, thus making it unattractive to deny cooperation. However, this mechanism also has limitations. First, although it enables the isolation of malicious nodes, it is vulnerable to false accusation, if trusted nodes lie. Second, in case the specific node does not exceed the predefined threshold,

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

which are used to distinguish deliberate malicious behavior, and move out of the range, which are shared by nodes of friend list, it needs much time to detect the moving malicious nodes. Third, alarm messages concerning detected malicious nodes are only sent to a friend list, which are previously registered, thus other node on this list cannot be recognized.

In summary, there are several security issues in detection and reaction mechanisms; first, the problem of how to accurately and quickly detect false accusations. Second, the problem of how to keep track of moving malicious nodes, which their weight values is not exceed to the predefined threshold. Finally, normal nodes should not be isolated due to the ambiguous collisions. So incorrectly imposed weight value for well-behaved nodes must be relieved. In this paper we present a solution of above problem.

Ⅲ. Assumptions and Background

A. Assumptions

Misbehaving nodes are a severe threat to the correct routing functionality in the ad hoc network. Before presenting our proposed scheme, we discuss the assumptions we made while designing the solution. For our scheme, we assume the following characteristics.

1) Selfish and malicious node: Misbehaving nodes are categorized by two types of nodes. A selfish node wants to preserve own resources while using the services of others. Otherwise, a malicious node that is not primarily concerned with power saving but that is interested in attacking the network in order to breakdown or partition into the network.
2) Promiscuous mode operation: We assume wireless interfaces that support promiscuous mode operation. This means that if node A is within range of node B, it can overhear communications to and from B even if those communications do not directly involve A.
3) No colluding nodes: Since this mechanism aim at the special circumstance, we assume there is no colluding between neighboring nodes in a path from source to destination.
4) As using region key sharing between upper layer and lower layer, it is protected overlay regions. And as using pair-wise key between neighboring nodes, it can authenticate each other.

B. Background

- Multipath establishment procedure(based on AODV)

To quickly send the information at the destination node without delay time, we propose the establishment of multiple paths less than four while detecting the routing path. If routing protocols can discover multiple paths, it can easily switch to an alternative path when the primary path appears to have failed.

Proposed multipath establishment procedure basically follows those of AODV[16]. However, to establish multipath, the destination node not unicasts a route reply (RREP) packet but sends the RREP to each neighbor forward a route request (RREQ) packet. Using the multipath is more useful than a single path because of not needing to rebroadcast RREQs for another path discovery when the primary path appears to have failed. And also traffic overheads for multipath establishment are less than those of broadcasting RREQs for path discovery because multipath establishment only temporally requires the memory capacity of several nodes.

Multiple path establishment procedure is done in following steps.

1) The source node initiates the path discovery by broadcasting the RREQ to its neighbors when the source node needs to communicate with another node for which it has no routing information in its table. The RREQ contains the following fields : <*source_addr, source_sequence_#, broadcast_id, dest_addr, dest_sequence_#, hop_count*>.
2) Each neighbor rebroadcasts the RREQ to its own neighbors after increasing the hop_count until reach the destination node. If an intermediate node has already established the path information for the desired destination in a route table, it drops the RREQ and unicasts the RREP back to its neighbor from which it received the RREQ.
3) Eventually, the RREQs will arrive at the destination node through various paths. As the RREQs travel from the source to the destination, it automatically sets up *the reverse paths* from all nodes back to the source. The destination node unicasts the RREP back to each neighbor which forwards the RREQs until the third arrival order. As the RREPs travel to the source, the primary and alternate paths are constructed.

Ⅳ. Detection and Management of False Accusations

As aforementioned, there have been several mechanisms to detect misbehaving nodes. However, these studies show that the network is still vulnerable if misbehaving nodes make a maliciously false accusation about normal nodes. This is a serious problem owing to the possibility of isolating normal nodes from the network. To solve this problem, we propose a new efficient mechanism for detecting and managing false accusations of malicious nodes.

A. Threat Model

There has been no research on efficient mechanisms to detect false accusations. Let us consider the scenario presented in Figure 1. Even though node C correctly

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

forwards packets to destination node D, to isolate node C from the network, malicious node B send alarm(B,C) message to source node S. Alarm(B,C) denotes that node B confirms that node C is a malicious node. When destination node D receive packets from source node S, destination node D unicasts an ACK message to source node S. However, malicious node B discards the ACK message to conceal his behavior. If source node S does not receive the ACK message during the constant period, he resends data and alarm(B,C) message through alternative path (S-I-J-K-D). Finally, the normal node C could be isolated by the false accusation of malicious node B.

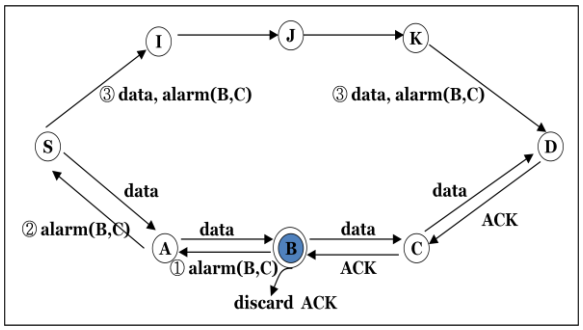


Fig. 1. Example of the false accusation

B. Detection mechanism of the false accusation

As mentioned in 4.1 Section, to detect false accusation nodes we propose a new efficient detection mechanism. Each node maintains a *Suspect Node List (SNL)* to manage malicious nodes and an *Isolate Node List (INL)* to isolate identified malicious nodes. As you can see in figure 2, let us consider that node B maliciously sends alarm(B,C) message to source node S to isolate the normal node C. If source node S receives alarm(B,C) message without receiving ACK message from destination node, he resends the data and alarm(B,C) message through alternative path(S-I-J-K-D). Once receiving duplicate the same data from the source node, destination node D can identify that alarm(B,C) message received from node B is the false accusation. Therefore, destination node D unicasts the RREP(B) message meaning malicious node B to source node S.

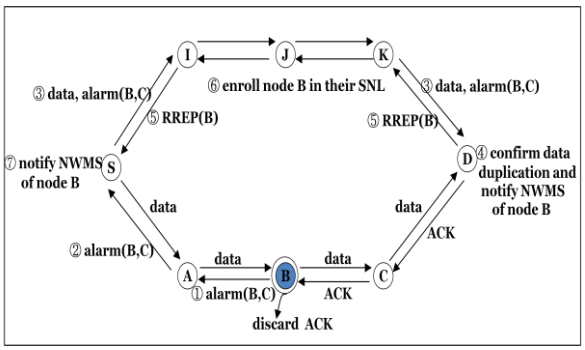


Fig. 2. Detection mechanism of the false accusation

Intermediate nodes (K, J, I) enroll node B in their SNL. If source node S receives the RREP(B) message, he can know that node B is malicious. And then source node S notifies the NWMS of that information.

However, even though source node S and destination node D notify the NWMS of malicious node B, it is not immediately valid. The NWMS determines whether the notice is true by comparing the information received from both source node and destination node during the constant period. Figure 3 represents working principle of the NWMS. The NWMS maintains *Misbehaving Node List (MNL)* to keep track of every misbehaving node in their area. Once the information of the malicious node is received from the source and destination node, and if both messages are same, the NWMS verifies whether the notified node is included in the MNL. If on the MNL, the malicious node' weight is increased by one. If not on the MNL, it will be enrolled on the MNL and assigned an initial value of weight of one. If the weight of node enrolled in the MNL exceeds the threshold, the NWMS broadcasts the fact that the node confirmed as malicious should be isolated to every node in their area. Every node received the information enrolls the malicious node in their *INL*, the malicious node is isolated from the network.

C. Detection mechanism of the selfish node

Our mechanism can also effectively detect selfish nodes. Figure 4 represents the detection mechanism of a selfish node. If destination node D confirms the data received from source node S whether duplicated, he can easily determine whether alarm (A,B) is malicious.

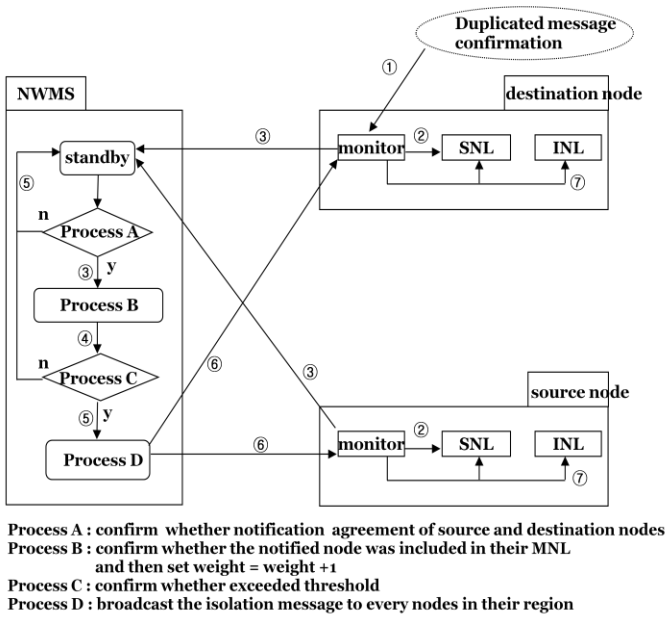


Fig. 3. Working principle of the NWMS

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

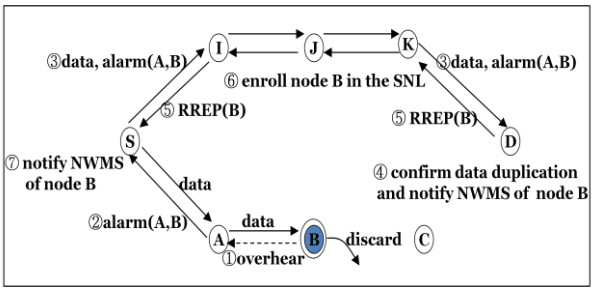


Fig. 4. Detection mechanism of the selfish node

D. Rescue mechanism of incorrectly imposed weight value

If a normal node was incorrectly imposed weight due to ambiguous network collision, rescue method is needed. To rescue normal nodes, each node maintains a SNL and monitors suspected neighboring nodes. In other words, if each node detects the normal behavior of neighboring nodes in their SNL, they reduce weight of that node. The SNL is not used to set routing path but only used to manage suspected nodes. Figure 5 illustrates the rescue procedure of incorrectly imposed weight value on a normal node in the NWMS. If node A detects normal behavior of node B in its' SNL, it reduces one with weight value of node B and then reports it to the NWMS. The NWMS checks whether weight value is zero in the MNL, if node B has weight value of zero then the NWMS deletes node B from the MNL and broadcasts a message to erase suspected node B from the SNL to every node in their region, otherwise the weight is reduce by 0.1. All nodes receiving the message immediately erase node B from their SNL.

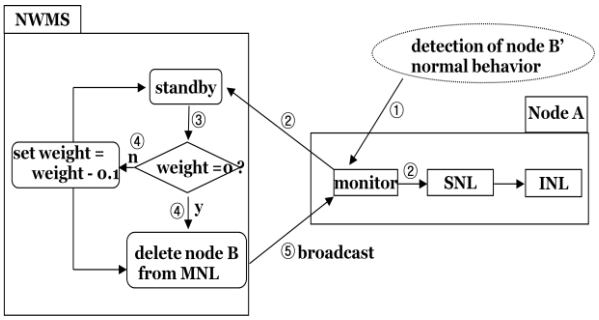


Fig. 5. Rescue procedure of the incorrectly imposed weight value in the NWMS

E. Node inner procedure steps from outer event

Figure 6 illustrates procedure steps when intermediate nodes receive a RREP message including a suspect node coming from a destination node. Once a node receives this message, he checks whether the suspected node was included in the SNL. If it isn't then enroll in the SNL and set count with 5, otherwise increase count by 5(maximum=10).

Figure 7 illustrates inner procedure steps of each node to alleviate the count when detecting normal behavior of neighboring nodes in their SNL. Once the nodes detect normal behavior of neighbor node in their SNL, they notify the NWMS of the information and reduce the count by one. If the count reaches zero then erase that node from the SNL.

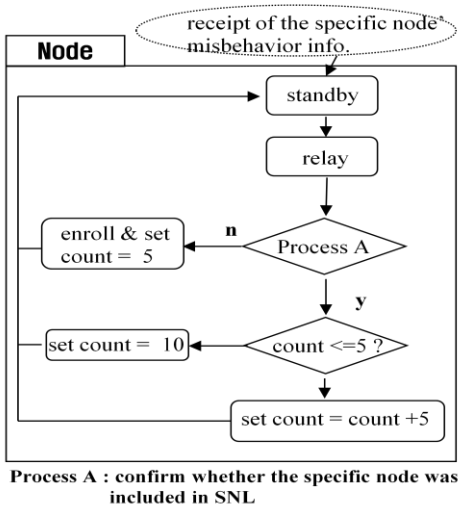


Fig. 6. Procedure steps when intermediate nodes receive a RREP message

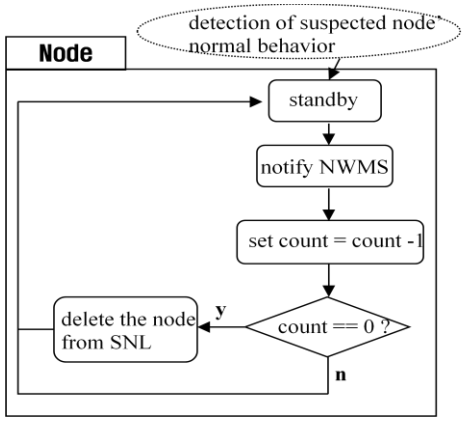


Fig. 7. Procedure steps when each node detects normal behavior of the node in their SNL

V. Simulation and Results

In order to evaluate the quality of our mechanism, we perform a simulation by comparing with AODV.

A. Simulation environment and scenario

In this simulation, to reduce the communication overhead, the *unicast* mode is used for every report management packets between nodes; the broadcast mode is only used for the NWMS alarms misbehaving nodes to every node. The list of simulation parameters for our study is shown in Table

1. For our simulation, we use Network Simulator 2 (NS-2) version 2.1b9a. We set the weight increasing rate of the NWMS to '1', threshold to '5', the count increasing rate of the SNL in each node to '5', decreasing rate to '5' and maximum value to '5'. These are the optimal settings found as a result of running many simulations.

Table1. Simulation parameters

Parameter	Level
Area	1000m X 1000m
Simulation time	1000sec
Packet generation interval	100ms
Packet size	64 byte
Number of nodes	250
Speed	0 ~ 5m/s
Threshold	5
Radio Range	250m

B. Results

- Packet throughput as changing the number of misbehaving nodes

Figure 8 shows packet throughput of AODV versus our mechanism as varying the number of misbehaving nodes such as 25, 50. Since we set packet generation interval to 100ms, the number of generated packet per 100 second is 1000. We set the pause time to each 100 second and evaluate the packet throughput for each mechanism. The figure demonstrates that our mechanism can process packets over 90% after passing 200 seconds for each case, but AODV only process packets about 40~60%. The main reason is because our mechanism can efficiently detect and isolate the detected misbehaving nodes from the network in a timely manner.

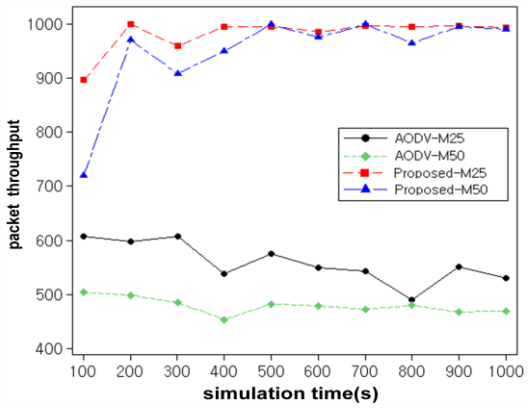


Fig. 8. Packet throughput as varying the number of misbehaving nodes

- The number of packet loss as increasing the number of nodes

Figure 9 shows the number of packet loss as increasing the number of nodes in case of the containment ratio of misbehaving nodes such as 10, 30%. Simulation time is 30 minutes. Figure demonstrates that packet loss of proposed mechanism is less than that of AODV. This is mainly because AODV do not handle those nodes when containment ratio of misbehaving nodes rises according to increasing the number of node. However proposed mechanism can detect and isolate misbehaving nodes in a timely manner. Therefore, the number of packet loss is kept on reasonable level.

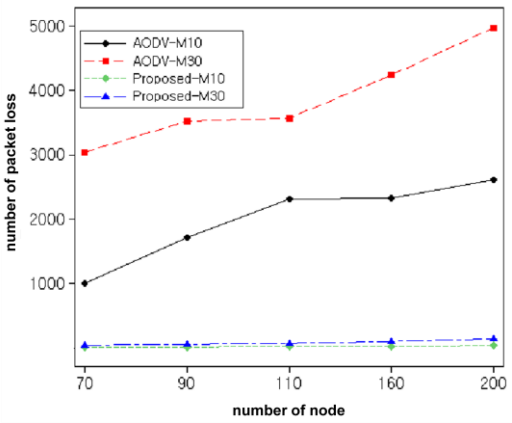


Fig. 91. The number of packet loss as increasing the number of nodes

- The number of packet loss as increasing the containment ratio of misbehaving nodes

Figure 10 shows the number of packet loss of AODV versus our mechanism as increasing the containment ratio of misbehaving nodes. Simulation time is 1000 seconds. Since AODV have not the method to handle the misbehaving node, packet loss ratio reaches 98% in case the containment ratio of misbehaving nodes exceeds 30%. However, in our mechanism, the number of packet loss reasonably increases because the proposed mechanism can quickly detect and isolate misbehaving nodes in the network. This results show that our mechanism handles misbehaving nodes efficiently. However, since the number of normal node is sparse as increasing misbehaving nodes in the network, the performance of our mechanism significantly depreciates when containment ratio exceed over 80%.

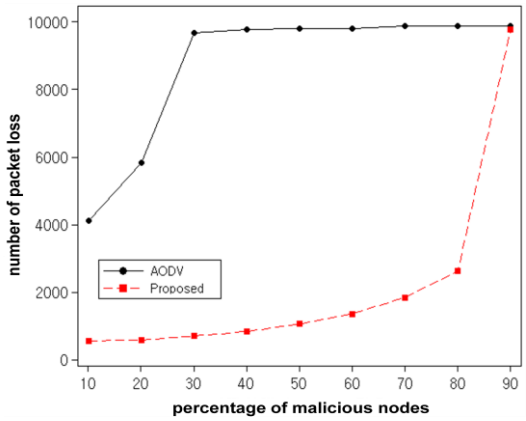


Fig. 20. The number of packet loss as increasing the containment ratio of misbehaving nodes

- **Routing overhead**

The proposed mechanism needs to more communication packets than the existing studies because of generating control packets to detect and isolate misbehaving nodes between the NWMS and nodes. Additional generated packets are as follow: each node reports misbehaving nodes to the NWMS defined as *U* and the NWMS broadcasts the determined misbehaving node to the every node defined as *B*.

In case of the number of packet generation per second *t*, the average number of node on routing path *p*, packet size NP_{size}, total number of nodes *N*, the number of misbehaving nodes *M*, packet size to alarm and report CP_{size}, simulation time *T*, the average number of node between the NWMS and node *n*,

$$U = n \times CP_{size} \times \text{threshold} \times M$$

$$B = M \times CP_{size} \times N$$

Therefore, total routing overhead (O) is denoted by

$$O = \frac{U + B}{T \times t \times p \times NPsize}$$

In this simulation, extra routing overhead rises about 0.97%. However, using our mechanism increases the throughput of the network by about twice (see Figure 8). Therefore, this routing overhead is negligible compared to the effect of the packet throughput.

Ⅵ. Conclusion

Ad hoc networks have vulnerabilities according to relying on the cooperation of all the participating nodes. Therefore, existing secure mechanism for ad hoc network do not efficiently detect the false accusations of malicious nodes.

In generally, the tactical ad hoc network does not correspond with the original ad hoc network as following potential features: a) initial nodes can be entirely trusted, b) some of nodes can be the upper level node which have enough resource and computing power. Therefore, in this paper, we employed upper level node as the NWMS to efficiently detect and isolate misbehaving nodes and also applied the alleviation mechanism for the weight value which was falsely imposed to prolong lifetime of nodes. In particular, we have proposed an efficient mechanism to resolve the false accusation problem caused by malicious nodes intentionally pointing out a normal node as a malicious node.

As confirmed from the simulation results, our mechanism has proved to be reliable. Using this mechanism we can significantly improve the management performance for misbehaving nodes.

References

[1] S. Zhu, S. XU, S. Setia, and S. Jajodia, "Establishing Pairwise Keys for Secure Communication in Ad hoc Networks: A Probabilistic Approach," Proceedings of the 11th IEEE Conference on Network Protocols (ICNP'03), pp.326-335, Nov. 2003

[2] M. G. Zapata, and N. Asokan, "Securing Ad hoc Routing Protocols," In Proceeding of the 2002 ACM workshop on wireless security, pp.1-10, Sep. 2002

[3] N. AsoKan, and P. Ginzboorg, "Key agreement in ad hoc networks." Computer Communication. pp.1627-1637, 2000

[4] L. Zhou, and Z. J. Haas, "Securing Ad Hoc Networks," IEEE Network Magazine, 1999

[5] C. K. Toh, "Ad Hoc Mobile Wireless Networks: Protocol and Systems," Prentice Hall PRT, 2002

[6] Frank Stajano and Ross Anderson. "The Resurrecting Duckling," Lecture Notes in Computer Science, Springer-Verlag, 1999.

[7] Panagiotis Papadimitratos and Zygmunt J. Hass. "Secure Routing for Mobile Ad Hoc Networks," In SCS Communication Networks and Distributed Systems Modeling and Simulation Conference (CNDS 2002), San Antonio, TX, Jan. 2002.

[8] Y. Hu, A. Perring, and D. Johnson, "Ariadne: A Secure On-Demand Routing Protocol for Ad Hoc Network," in proceedings of ACM MobiCom '02, 2002.

[9] Dave B. Johnson and David A. Maltz. "The dynamic source routing protocol for mobile ad hoc networks," Internet Draft, Mobile Ad Hoc Network (MANET) Working Group, IETF, Oct. 1999.

[10] K. Sanzgiri, B. Dahill, B. Levine, and E. Belding-Royer, "A Secure Routing Protocol for Ad Hoc Networks," in Proceedings of IEEE ICNP '02, 2002.

[11] F. Kargl and A. Geiss, "Secure Dynamic Source Routing," in Proceedings of HICSS 38. 2005

[12] S. Marti, T. J. Giuli, K. Lai, and M. Baker, "Mitigating routing misbehavior in mobile ad hoc networks," Proceeding of the 6th International Conference on Mobile Computing and Networking, pp.255-265, Aug. 2000.

[13] S. Buchegger, and J. L. Boudec, "Performance analysis of the CONFIDANT Protocol," Proceeding of the 3rd ACM International Symposium on Mobile ad hoc networking & computing, pp.226-236, 2002

[14] C. K. Toh, C. Lee, and N. A. Ramos, "Next-Generation Tactical Ad Hoc Mobile Wireless Networks," Technology Review Journal Spring, pp.103-113, 2002

[15] J. Brand, and G. Hart wig, "Management of tactical ad hoc networks with C2 data models," Military Communication Conference 2001 IEEE, pp915-922, Aug. 2002

[16] C. E. Perkins, E. M. Royer, and S. R. Das. "Ad hoc on-demand distance Vector (AODV) routing," IETF, MANET working group, Oct. 1999

AUTHORS PROFILE

Yunho Lee received his B.S. in Electronic Engineering from Korea Military Academy in 1999, M.S. in Computer Engineering from Seoul National University, Korea, in 2005. Currently, he is a Ph.D. course in Computer Science, Korea National Defense University. His research interests include mobile network security, and intrusion detection.

Soojin Lee received his B.S. in Computer Science from Korea Military Academy in 1992, M.S. in Computer Science from Younsei University, Korea, in 1996, and Ph.D. in Computer Science from Korea Advanced Institute of Science and Technology (KAIST), Korea in 2006. Since 2006, he has been an associate professor at the Dept. of Computer Science, Korea National Defense University. His research interest includes computer and communication security, intrusion detection, and mobile network security.

Wavelength Requirements for a Scalable Single-hop WDM Optical Network

Rabi W Yousif and Borhanuddin Mohd Ali
Faculty of Engineering, Universiti Putra Malaysia
Selangor, Malaysia
rabi.habash@gmail.com

Mohd Khazani Abdullah
Significant Technologies Sdn. Bhd.
Serdang, Malaysia

Kamaruzzaman Bin Seman
Faculty of Science and Technology
Universiti Sains Islam Malaysia, Nilai, Malaysia

Mohd Dani Baba
Faculty of Electrical Engineering
University Teknologi Mara, Shah Alam, Malaysia

***Abstract*—In this paper, we present a method for designing a passive optical based single-hop wavelength division multiplexing multicast architecture that can achieve a scalable structure and form the basis of a wavelength efficient single-hop WDM network. The proposed architecture minimizes the number of wavelengths required for efficient multicast service and also minimizes tunability requirement of the transceivers. The network size scalability is achieved by adding transmitters and receivers to the designated groups. We show that the proposed system can accommodate large tuning delays and keeps with suitable throughput when the number of wavelength is equal to the number of nodes. We also show that the design can lead to a scalable structure while minimizing the number of wavelengths and tunability of the transceivers required for an efficient multicast service resulting in an improved system throughput and delay performance.**

Keywords-multicasting, wavelength-division multiplexing, single-hop passive optical network

I. INTRODUCTION

In recent years, the Internet traffic has increased tremendously, because multimedia traffic such as video streaming service, high resolution images, digital video and audio conferencing, and business data distribution becomes prevalent in the Internet. Some multimedia applications require strict quality-of-service (QoS) or multicasting.

Current state-of-the-art dense WDM systems are using narrow 50-GHz (0.4 nm) channel spacing. In such systems, functions traditionally performed by electronics, such as switching, signal amplification, etc, are performed in the optical domain, therefore achieve signal transparency. Thus, the capability for multicast transmission has become a very important requirement for access networks [1, 2].

WDM technology has the potential to satisfy the ever-increasing bandwidth needs of network users on a sustained basis. Today, optical backbones with a transmission speed of 40 gigabits per second are deployed. This technology is reliable and will meet bandwidth needs for the next few years. However, considering that traffic is growing by 40 percent a year on average, even 40G networks will have to be expanded to 100G. WDM optical networks can efficiently support multicasting since splitting light is inherently easier than copying data into an electronic buffer. Applications of multicasting include multimedia conferencing, distance education, video distribution, distributed games and many others [3, 4]. For cost reasons each node in single–hop WDM networks deploys a rather small number of transceivers which is typically smaller than the number of wavelengths available for data transmission/reception. To increase the network efficiency all wavelengths should be used at any given time.

In single hop communication, the network must be able to establish any possible connection in one hop, without intermediate relaying or routing. This in effect implies that the network will have to change the connections it supports at different times. Multi-hop networks have the ability to circumvent the network capacity limitations. Each node is connected to only a few other nodes, as such only few wavelengths are required per node. This greatly reduces the wavelength bottleneck.

II. KEY DESIGN REQUIREMENTS

When designing a WDM network architecture and protocol, the following key requirements and properties have to be satisfied [5 - 8]:

- Provide point–to–multipoint connections in order to support multicast applications such as videoconferences and distributed games in an economical and bandwidth–efficient manner.
- Add or remove network nodes in an easy and nondisruptive way without significantly degrading the network performance.
- Traffic should not have to traverse a large number of intermediate nodes to ensure smaller resource requirements and smaller propagation delays.
- Provide some level of assurance that the service requirements for different types of traffic, e.g., for delay–sensitive, real–time, and interactive applications, are satisfied.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

- Allocate network resources to all nodes which need to send data. In networks with fair channel access control each node ready to send data should have an equal opportunity to transmit.
- To cope with the resulting increased local traffic, metro networks have to be easily upgradeable. Advanced technologies, e.g., tunable transceivers with a wider tuning range and a smaller tuning time, have to be incorporated without network service disruption and reconfiguration.

III. System Description

The system under study is based on a broadcast-and-select WDM architecture consisting of N network nodes connected via optical fibers to a passive star coupler (PSC) as shown in Figure 1. There are W wavelength channels, where $W \leq N$. The bandwidth of a fiber is divided into $W+1$ channels, where $W \leq N$. One of the channels, λ_0, is used as a control channel which is shared by all nodes. The rest of the channels, $\lambda_1,\ldots,\lambda_W$, are used as data channels.

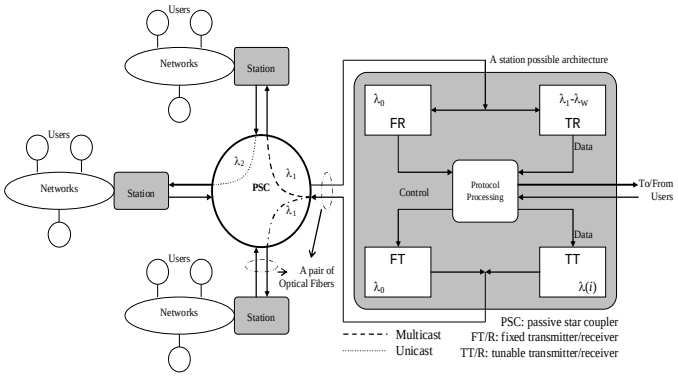


Figure 1. A broadcast-and-select star-based WDM optical system.

Each node in the network is connected to the PSC by a transmitting and receiving fiber, and each message is addressed (multicast) to a number of receivers (destination set size), randomly chosen from the N network nodes and each receiver tunes to one of the wavelengths that has a message addressed to it. Also, each node has one fixed transmitter and one fixed receiver in order to access the control channel. Moreover, in order to access data channels, each node has one tunable transmitter and one tunable receiver, so that full connectivity can be achieved by tuning transmitters to the different wavelengths.

Tuning times are not negligible with respect to the slot time. A centralized network controller allocates slots in a WDM frame according to (long-term) bandwidth requests issued by users. When $W \leq N$, two or more nodes share one data channel. Each node is equipped with a buffer in which arriving data packets are stored. Deploying tunable transmitters and receivers at each node allows for load balancing since traffic between a given pair of nodes can be sent on any wavelength. In particular for nonuniform traffic, load balancing increases the channel utilization and improves the throughput–delay performance of the network.

All stations can communicate with one another. In addition, a pair of fixed transceivers and control receiver both are tuned to the control channel is dedicated for pre-transmission co-ordination. However, communication between two nodes is possible only when the transmitter of the source node and receiver of the destination node are tuned to the same channel during the period of information transfer.

IV. System Assumption

The behavior of the system is characterized by the following assumptions:

- There are N network nodes and W wavelength channels in the system.
- Each node has a single–packet buffer, i.e., each node can store at most one data packet at any given time.
- After transmitting a data packet in a given frame the buffer becomes empty at the end of that frame.
- Each message is multicast to a set of l receivers where $l < W \leq N$.
- Whenever the receivers of a multicast group are ready to receive a data packet the source node's transmitter is ready to transmit.
- A packet that arrives at the start of a slot can be transmitted during that slot to any one of the other (N – 1) nodes with equal probability.
- A node sends out its control packet in a frame with probability p, not only for retransmissions but also for first–time transmissions.
- Random selection of a destination node among the (N – 1) nodes is renewed for each attempt of transmitting a control packet.

V. System Model

A. Node Structure

The proposed architecture aims to define a minimum group of network nodes for a local structure, assign a unique wavelength to a transmitter, and identify, for each transmitter, the minimum set of additional wavelengths needed to achieve communication with every other node in the local cluster and hence all the nodes in the network.

Figure 2 shows the node structure of the system. Each receiver is able to tune to all the wavelengths assigned to the transmitters having direct links to it. Each processor can transmit data on a fixed number of wavelengths, but can receive data on a range of wavelengths by dynamically tuning to the wavelength of a transmitting station. All the processors are synchronized at the optical coupler. The use of the same structure for both the transmitter and receiver is strategic [6 - 9]. This will greatly simplify the coupling of the local structure. Each node can switch channels (wavelengths) during execution by dynamically changing the injection current to the laser.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

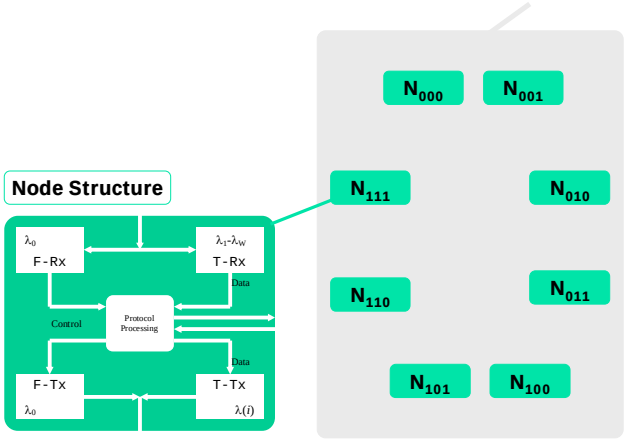


Figure 2. A network node structure.

Additionally, transmission and reception can be performed on different channels. The single star topology consists of *n* inputs, to which one transmitter is connected, and *n* outputs, to which one receiver is connected.

B. *Connection Establishment and Partitioning*

There are basically two ways to achieve connections between nodes in an optical network, path multiplexing and link multiplexing. In the first the same wavelength has to be assigned all to the links between source and destination, while in the second, different wavelengths can be assigned on different links. To achieve single-hop connectivity, a wavelength allocation mechanism needed to determine a path for a new request. Each transmitter group can have direct links to exactly two receiver groups. For any transmitter group, the two receiver groups that do not have a direct link to it consist of those that contain receivers with the same index notation as one of the transmitters in the transmitter group. For any transmitter/receiver group, there are two receiver/transmitter groups that can have direct links to it and two others that do not have direct links to it [9]. The transmitter/receiver groups not having direct links to the same receiver/transmitter groups are mutually exclusive. Finally, half of the number of transmitters/receivers can be directly connected to half of the number of receivers/transmitters simultaneously. Figure 3 shows the connection establishment procedures among the network nodes in order to achieve single-hop communication.

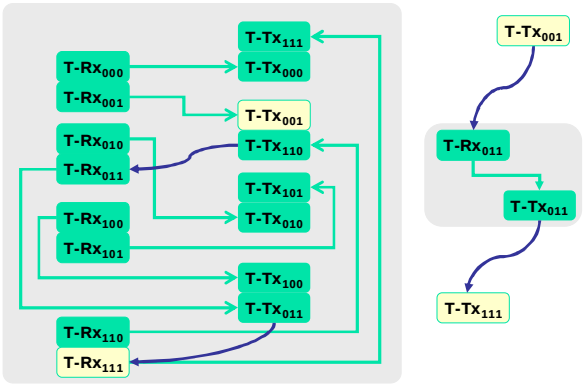


Figure 3. Connection establishment procedures among the nodes.

Partitioning is desirable to design a hierarchical structure using a cluster-based approach. This becomes quite obvious since it is intended that the network should be scalable, flexible and implemented with wavelength division multiplexing techniques. Each local unit should be autonomous, so that wavelength reuse can be achieved.

C. *Scalable Approach*

If there are more than *m* access nodes, where *m* is the desired number of access nodes representing the regular local structure, a partition can then be accomplished by defining a minimum set of access nodes as the local structure and applying the partition mechanism that is explained above to achieve the partition set. The transmitter in a group needs to be placed according to the partitioning mechanism and same also for the receiver. Scalability here has two aspects. First, the transmitter and receiver of the new access node must be physically connected to the optical medium and second, the added access node must be incorporated into the MAC mechanism that controls the single hop connections [8-10].

To incorporate the added access nodes in the MAC mechanism, it requires only modifications for the control channel. This means that the number of added nodes must increase the number of control slots. To correctly reach each added node, all transmitters must be informed about its receiver configuration and its address.

D. *Wavelength Allocation*

We assign wavelengths such that the tunability for the transmitters is minimum and optimal while the receivers must be able to tune to the maximum number of wavelengths used in the entire network. Higher number of tuning disrupts the network due to the retuning time. Our goal is to minimize the number of tuning so that the reconfiguration does not suspend the operation of the network.

Since computer networks traffic changes rapidly, there is a need for a good mechanism to change the current situation of the network in terms of wavelength allocation (i.e., the current wavelength assignment into a new wavelength assignment). However, the number of channels is a limiting factor in a WDM network, and is typically less than the number of nodes in the network. Therefore, more than one receiver is assigned to one channel. This problem is called wavelength assignment problem.

Wavelength blocking is a major problem with path multiplexing. One obvious disadvantage with the link multiplexing is the use of wavelength converters at intermediate nodes to eliminate blocking. This however, increases the cost and complexity of the system.

In single-hop communication, all the nodes can reach any other node directly. This means that the transmitted data are not passed through any intermediate routing stages and remain in optical form all the way from the source node to the destination node. In such mode of communication, a lightpath is established before a communication starts and the data transmission is carried out in a pure circuit-switched manner.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

With dynamic traffic demands, new lightpaths need to be added to the logical topology each time an arriving connection request cannot be accommodated. The WDM techniques enable extraction of a larger amount of usable bandwidth. Routing and wavelength assignment algorithms fine-tune the overall process by achieving orders of magnitude of performance improvement. The goal is to present an efficient dynamic distributed routing and wavelength allocation method that minimizes path latency, wavelength blocking and the number of wavelengths applied [12-13].

E. Multicast Scheduling

The multicast scheduling problem can be described as follow: given N stations, W available wavelengths for data transmission, L slots global cycle and a $W \times L$ slots allocation matrix D, each station is equipped with a pair of tunable transceiver and each needs δ time slots for tuning from λ_i to λ_j, $i \neq j$.

For a setup request, find a new feasible slot allocation matrix D_{new} with a new global cycle length L_{new} such that setup request is arranged into D_{new} and all the QoS requirements of accepted multicast in D are not affected.

The proposed system follows a broadcast-and-select methodology. It is assumed that each processor has a fixed-tuned transmitter and a fixed-tuned receiver for transmitting and receiving on the control channel. In addition, each processor has a tunable wavelength transmitter and a tunable wavelength receiver for transmitting and receiving on the data channel. Each processor can transmit data on a fixed number of wavelengths, but can receive data on a range of wavelengths by dynamically tuning to the wavelength of a transmitting station. The broadcast function is achieved using a free-space optical coupler, while the select function is achieved by using the wavelength tunable photodetector with filters. All the processors are synchronized at the optical coupler. The transmitted signals from each transmitter are routed to a free-space optical star coupler via optical waveguides. The signals are also routed back to the receivers via optical waveguides.

When the traffic demand changes in a network, the current wavelength assignment becomes unbalanced and requires to be changed. A new wavelength assignment is calculated depending on the new traffic demand and a different wavelength assignment will be produced, which requires a number of tunings. However, if the current wavelength allocation is taken into account when calculating the new one, the required number of tunings can be reduced with an acceptable wavelength assignment [4, 11].

The traffic in the network is represented by a traffic matrix $T = \{t_{ij}\}$ where t_{ij} is the rate of traffic from node i to node j where $i, j = 1,\ldots, N$. From the traffic matrix the total bandwidth requirement b_j of receiver j can be computed, which equals the sum of the elements of the jth column of T:

$$b_j = \sum_{i=1}^{N} t_{ij} \tag{1}$$

Now assume that R and R' represent a current wavelength assignment and a new wavelength assignment respectively. Based on b_j for the N receivers where $j = 1,\ldots, N$, the received traffic for each node can be calculated. The wavelength assignments R and R' refers to the association between the nodes and the wavelengths, then

$$R = \{R_c , w = 1,\ldots, W\} \tag{2}$$

where R_c is the set of receivers that are assigned to wavelength w, $w = 1,\ldots, W$ in the current wavelength assignment, i.e.,

$$R_w = \{j \mid \lambda(j) = \lambda_w\} \; w = 1,\ldots, W \tag{3}$$

where $\lambda(j)$ is the channel to which the receiver j is assigned. For example, if $\lambda(3) = \lambda_2$, it means that node 3 receiver is assigned to channel 2. And

$$R' = \{R'_w, w = 1,\ldots, W\} \tag{4}$$

where R'_w is the set of receivers that are assigned to w in the new wavelength assignment, i.e.,

$$R'_w = \{j \mid \lambda(j) = \lambda'_w\} \tag{5}$$

Now, assume number of tunings is $D(R, R')$, i.e., the number of receivers that need to be retune to take the network from its current wavelength assignment R to the new wavelength assignment R' and can be given by

$$D(R, R') = N - \sum_{w=1}^{W} \left| R_w \cap R'_w \right| \tag{6}$$

Figure 4 shows the current wavelength assiginment for a network with 6 network nodes and 3 wavelength channels, then $R = \{R_1, R_2, R_3\}$ where $R_1 = \{4, 5\}$, $R_2 = \{1\}$ and $R_3 = \{2, 3, 6\}$.

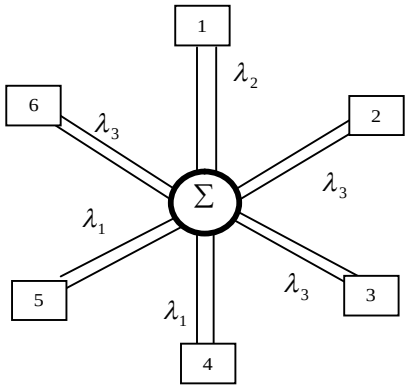


Figure 4. Current wavelength allocation for a network with 3 wavelengths and 6 nodes.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

When the traffic demand of the network changes, the wavelength assignment is updated as shown in Figure 5.

$R' = \{ R'_1, R'_2, R'_3\}$ where $R'_1 = \{4, 5\}$, $R'_2 = \{3\}$ and,

$R'_3 = \{1, 2, 6\}$. Then $R_1 = R'_1$

$R_2 \neq R'_2$, $R_3 \neq R'_3$

Since $D(R, R') = N - \sum_{w=1}^{W} \left| R_w \cap R'_w \right|$, therefore,

$D(R, R') = 6 - 4 = 2$.

As a result, two nodes in the network need to tune their receivers.

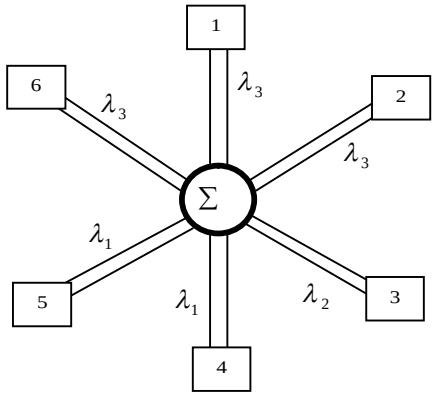


Figure 5. Updated wavelength allocation for a network with 3 wavelengths and 6 nodes.

VI. Results and Discussion

Figures 6 and Figure 7 show a trade-off between the required number of tunings and the number of wavelengths with and without partitioning for different network population. The tuning here is carried out to equalize as much as possible the busy channel and the free channel. The difference in number of tunings comes from the difference in the probability of exchanging two network nodes between two different loaded channels in which this probability decreases when the number of channels increases.

This is because when the number of channels increases, the number of receivers assigned to channels decreases. This in turns means that the maximum number of tunings is inversely proportional to the frequency of tuning, so that, the most important factor that affects number of tuning is how frequently it is applied.

Moreover, the number of channels affects the number of tunings, but in small scale. When the number of channels increases, the total number of tunings decreases because fewer receivers are assigned for the busy channels. Hence, there is a small chance that the nodes are exchanged between two different loaded channels.

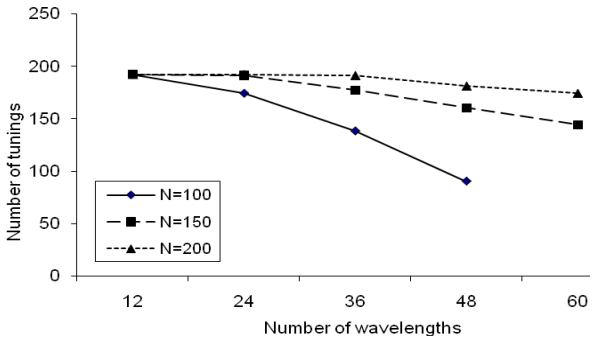


Figure 6. Number of tunings vs. number of wavelengths for a network with 50, 80, and 130 nodes and without partitioning.

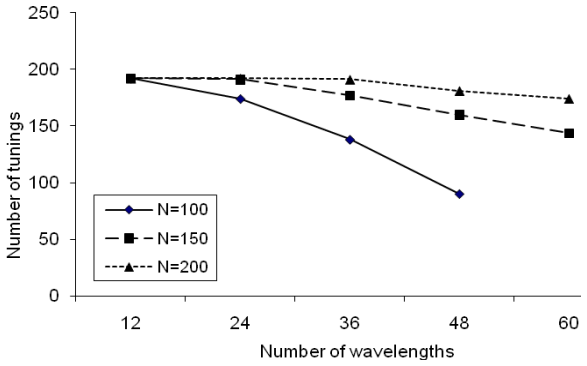


Figure 7. Number of tunings vs. number of wavelengths for a network with 100, 150, and 200 nodes with 4 partitions.

Figure 8 and Figure 9 show the effect of partitioning on the network performance in terms of the required number of wavelengths and number of tunings. It can be observed that the number of tunings decreases when the number of partitions increases. However, the performance is not affected by partitioning the network where the network performs almost the same for all cases and for most of the time the performance improves when number of partitions increases.

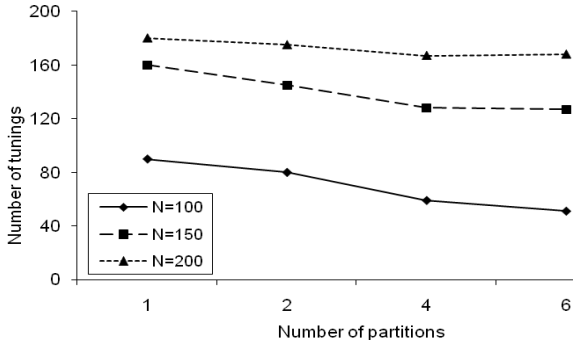


Figure 8. Number of tunings vs. number of partitions for a network with 100, 150, and 200 nodes and 50 wavelengths.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

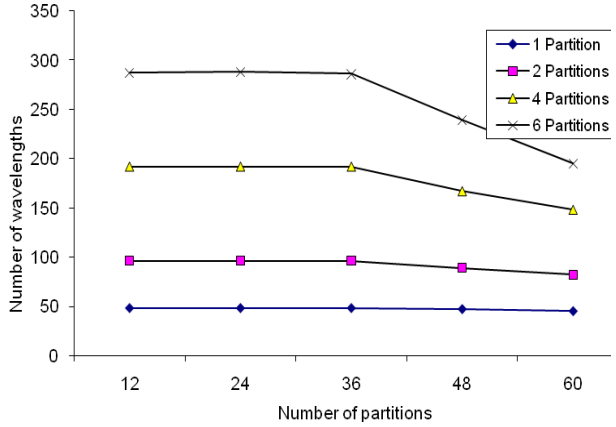


Figure 9. Number of wavelengths vs. number of partitions for a network with 200 nodes and 1, 2, 4, 6 partitions.

Figure 10 shows the average packet delay versus throughput characteristics of multicast, transmitter, and receiver for a PSC based single-hop WDM network with 200 network nodes and retransmission probability equals to 0.5. The network receivers are divided into two groups allowing wavelength reuse, which is possible during the reservation phase, i.e., the first slots of every frame when the control packets are transmitted.

As can be seen the transmitter throughput is affected by the many destination conflicting multicast transmissions. However, with two partitions multicast copies destined to the coupler will likely experience receiver conflicts since on average each copy is destined to more receivers for two partitions than for more partitions.

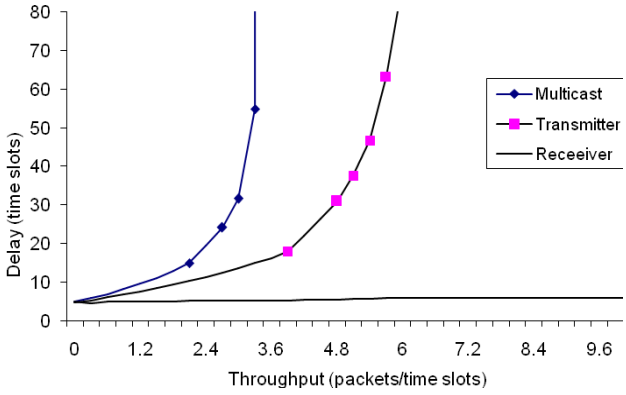


Figure 10. Delay vs. throughput (multicast, transmitter, and receiver) with 200 network nodes and, 2 partitions, and retransmission probability equals to 0.5.

VII. Concluding Remarks and Future Work

The proposed method of designing a PSC based single-hop WDM multicast architecture can achieve a scalable structure that can form the basis of a wavelength efficient single-hop WDM network. The proposed architecture minimizes the number of wavelengths required for efficient multicast service and also minimizes tunability requirement of the transceivers. The network size scalability is achieved by adding transmitters and receivers to the designated groups. Wavelength scalability is achieved through wavelength spatial re-use.

The problem of updating the wavelength assignment in single-hop WDM networks is considered where the traffic demand changes frequently and changing the channel assignment becomes necessary. Minimizing the number of tunings required can be achieved by exchanging one of the receivers, which is assigned to the channel with high load, with the appropriate receiver in the channel with minimum load. Tuning is carried out to equalize as much as possible the most loaded channel and the least loaded channel. In this environment, the problem is finding an allocation of wavelengths to receivers such that the number of transmissions of a multicast message is minimized. Since the number of wavelengths is limited by technology, the problem then becomes in finding the best partitioning scheme for the receivers in the network. The proposed system can accommodate large tuning delays and keeps with suitable throughput when the number of wavelength is equal to the number of nodes. When the number of wavelengths is comparable to the number of users the tuning time influence on the packet delay increases.

In the context of wavelength allocation, a study on increasing the number of exchanges by taking into account the channel which comes right after the most loaded channel can be considered for future work. Typically, this will include the optimization of the communication, analysis of the communication patterns, and connection scheduling and communication phase analysis. Also a study the impact of large tuning delays on reconfiguration process and on the network performance is required.

References

[1] Han, K.E. and Yang, W.H, “Design of AWG-based WDM-PON architecture with multicast capability”, In the Proceedings of the HSN’2008 Workshop, vol. 1, USA, 2007.

[2] Zhone Technologies. Zhone WDM Optical Transport Equipment Chosen to Extend Fiber Capacity and Increase Reliability for Pure Packet Network. Press Release, 2007. [online] : http://www.zhone.com/about/news/2007/on-telecoms-greece.en-us.n2n?print=1

[3] Huang, S., Dutta, R., and Rouskas, G.N, “Traffic grooming in path, star, and tree networks: complexity, bounds, and algorithms”, IEEE journal on Selected Areas in Communications, vol. 24, no. 4, 2006.

[4] Nina, S, “Multicast routing and wavelength assignment in WDM networks: a bin packing approach”, Optical Society of America Publications, 2006.

[5] Maier, M., Scheutzow, M., and Reisslein, M, “The arrayed-waveguide grating based single-hop WDM network: an architecture for efficient multicasting”, IEEE Journal on Selected Areas in Communications, 21(9), 2004.

[6] Okorafor, E., and Lu, M, “Efficient distributed control routing and wavelength assignment mechanism for a scalable hierarchical single-hop WDM all-optical interconnection network”, In Proc. 6th Joint Conf. on Information Sciences (JSIC), pp. 276 – 282, 2004.

[7] Hamad, A. Tao, W. Kamal, A., Somani, A.K., “Multicasting in wavelength-routing mesh networks”, Computer Networks 50 (16): 3105-3164, 2006.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[8] Rouskas, G.N. Hierarchical Traffic Grooming, Chapter Book, 2008.

[9] Rouskas, G.N., "Optical layer multicast: rationale, building blocks, and challenges", IEEE Network, vol. 17, no. 1, pp. 60-65, 2005.

[10] Pointurier, Y. and Brandt-Pearce, M, "Routing and wavelength assignment incorporating the effects of crosstalk enhancement by fiber nonlinearity", 2005 Conference on Information Sciences and Systems, The Johns Hopkins University, 2006.

[11] Alfouzan, I. Jayasumana, A., "Dynamic reconfiguration of wavelength-routed WDM networks", 26th Annual IEEE Conference on Local Computer Networks Proceedings, pp. 477 - 485, 2002.

[12] Banerjee, A., Park, Y., Clark, F., Song, H., Kim, K., and Mukherjee, B., "Wavelength-division-multiplexed passive optical network (WDM-PON) technologies for broadband access: a review", Journal of Optical Networking, vol. 4, no. 11, 2005.

[13] Luekijna, K. and Saivichit, C., "Multicast traffic reconfiguration in WDM network for single node failure design", In the Proceedings of ICACT'2007, pp. 1833-1838, 2007.

AUTHORS PROFILE

Rabi W. Yousif received the B.S. degree in Electronics and Communications Engineering from Mosul University, Mosul, Iraq in 1996, and the M.S. degree in Computer Systems Engineering from Universiti Putra Malaysia, Serdang, Malaysia in 2000. His current research interests include WDM optical networks architecture, protocols, and network security.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Enhanced Dynamic Source Routing Protocol using On Demand Passive Clustering

Shobha.K.R and Dr.K.Rajanikanth
M.S.Ramaiah Institute of Technology
Bangalore, India.
shobha_shankar@yahoo.com
principal@msrit.edu

Abstract— **In this paper an enhancement technique for Dynamic Source Routing (DSR) protocol using on demand passive clustering is proposed. DSR is a popular reactive protocol in Mobile Adhoc Networks (MANETs) which uses flooding for route discovery and route maintenance. Flooding can be easily restricted using clustering. Clustering restricts the set of forwarding nodes during flooding which in turn reduces the energy, cost and traffic overhead during routing in dynamic traffic and dynamic topology environment of MANET. Active clustering mechanisms require periodic refresh of neighborhood information and tend to introduce a large amount of communication maintenance overhead. This can be overcome using passive clustering which executes only when the user has data to transmit. Passive clustering dynamically partitions the network into clusters interconnected by gateways. Existing Passive Clustering techniques exploit the data packets for cluster formation and thus reduce the periodic exchange of explicit control packets. We have proposed an enhanced passive clustering technique which makes use of the control packets used for route discovery in DSR to establish a cluster. We have used the Lowest ID technique and first declaration wins technique to form clusters. We have analyzed the performance of this technique to reduce flooding by using single cluster head called primary cluster head. Then an analysis has been carried out with two cluster heads called primary and secondary cluster heads.**

Simulation results on DSR have shown that this enhanced passive clustering technique can reduce redundant flooding, with negligible overhead; - thus making DSR more efficient.

Keywords- MANETs; DSR; Flooding; Clustering; Passive clustering; Primary cluster head; Secondary cluster head;

I. INTRODUCTION

Mobile Adhoc Networks (MANETs) have recently been the subject of active research because of their unique advantages. MANETs are self-creating, self-organizing and self-administrating networks and do not require deployment of any kind of infrastructure. They offer special benefits and versatility for wide range of applications in military (e.g., battlefields, sensor networks etc.), commercial (e.g., distributed mobile computing, disaster discovery systems, etc.), and educational environments (e.g., conferences, conventions, etc.), where fixed infrastructure cannot be easily acquired. With the absence of pre-established infrastructure (e.g., no router, no access point, etc.), two nodes communicate with one another in a peer-to-peer fashion. Two nodes communicate directly if they are within transmission range of each other. Otherwise, nodes must communicate via a multihop route. To find such a multi-hop route, MANETs commonly employ on demand routing algorithms that use flooding or broadcast messages. Many ad hoc routing protocols [12] [13] [26] [27], multicast schemes [25], or service discovery programs depend on massive flooding. In flooding, a node transmits a message to all of its neighbors. The neighbors in turn relay to their neighbors and so on until the message has been propagated to the entire network. In this paper, we will refer to such flooding as blind flooding. As one can easily see, the performance of blind flooding is closely related to the average number of neighbors (neighbor degree) in the CSMA/CA network. As the neighbor degree gets higher, blind flooding suffers from the increase of (1) redundant and superfluous packets, (2) probability of collision, and (3) congestion of wireless medium [1]. Performance of blind flooding is severely impaired especially in large and dense networks [2]. When topology or neighborhood information is available, only subsets of neighbors are required to participate in flooding to guarantee the complete flooding. We call such flooding as efficient flooding. The characteristics of MANETs (e.g., node mobility, the limited bandwidth and resource), however, make the periodic collection of topology information difficult and costly (in terms of overhead). For this reason many on-demand adhoc routing schemes and service discovery protocols simply use blind flooding [12] [25]. In contrast with on-demand routing methods, the proactive adhoc routing schemes by virtue of periodic route table exchange, can gather topological information without much extra overhead. Thus, the leading MANET proactive adhoc routing schemes use route aggregation methods to forward routing packets through only a subset of the neighbors [26] [27].

In this paper, we focus on on-demand routing protocols and propose a mechanism for efficient flooding based on passive clustering. Existing passive clustering scheme [10] requires neither the deployment of GPS like systems nor explicit periodic control messages to identify the subset of forwarding neighbors. This scheme makes the following contributions compared with previous efficient flooding schemes (such as multipoint relay, neighbor coverage, etc): (1) it does not need any periodic messages. Instead, it exploits existing data packets by attaching few more extra fields; (2) it is very resource-efficient regardless of the degree of neighbor nodes or

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

the size of network. To our knowledge, passive clustering is the only scheme that provides scalability and practicality for choosing the minimal number of forwarding nodes in the presence of dynamic topology changes; (3) it does not introduce any startup latency; (4) it saves energy if there is no traffic; (5) it easily adapts to topology and available resource changes.

In this paper we propose an enhanced passive clustering scheme which uses route request packets (RREQ) instead of data packets for maintaining cluster. This technique reduces the flooding efficiently and also reduces the overhead in each and every data packet transmitted. We have simulated and tested this algorithm using DSR routing protocol.

The remainder of the article is organised as follows: In section 2 we discuss the various methods available for achieving efficient flooding. Section 3 gives an explanation about the algorithms we are proposing for achieving efficient flooding .Section 4 discusses the simulation parameters and results .The main conclusions for this paper are summarized in section 5.

II. RELATED WORK

Several papers [1] [6] [7] [8] have addressed the limitations of blind flooding and have proposed solutions to provide efficient flooding. However, because of the problem of finding a subset of dominant forwarding nodes in MANETs, all the work about efficient flooding has been directed to the development of efficient heuristics that select a sub-optimal dominant set with low forwarding overhead.

In [1] [6], the authors propose several heuristics to reduce rebroadcasts. More specifically, upon receiving a flood packet, a node decides whether to relay it or not based on one of the following heuristics: (1) rebroadcast with given probability; (2) rebroadcast if the number of received duplicate packets is less than a threshold; (3) distance-based scheme where the relative distance between hosts determines the rebroadcast decision; (4) location-based scheme where the decision is based on pre-acquired neighbor location information; (5) cluster-based scheme where only precomputed cluster heads and gateways rebroadcast. Our approach, passive clustering, differs from the above schemes in that it provides a more systematic method based on locally collected information (e.g., neighbor information, cluster states, etc.). Each node participates in flooding based on its role or state in the dynamically constructed cluster architecture instead of depending on local heuristics or on pre-computed clusters.

Another approach to efficient flooding is to exploit topological information [6] [7] [8] [24] [19]. In the absence of preexisting infrastructure, all the above schemes use a periodic hello message exchange method to collect topological information. Our approach does not require periodic control messages. Rather, it exploits on-going Route REQuest (RREQ) packets to exchange cluster-related information. The authors of [8] suggest two schemes called self-pruning and dominant pruning. Self pruning is similar to the neighbor-coverage scheme in [6]. With self-pruning scheme, each forwarding node piggybacks the list of its neighbors on outgoing packet. A node rebroadcasts (becomes a forwarding node) only when it has neighbors that are not covered by its forwarding nodes. While the self-pruning heuristic utilizes information of directly connected neighbors only, the dominant-pruning heuristic extends the propagation of neighbor information two-hop away. The dominant pruning scheme is actually similar to Multipoint Relay scheme (MPR) [7]. In Multipoint Relay scheme, a node periodically exchanges the list of adjacent nodes with its neighbors so that each node can collect the information of two-hop away neighbors. Each node, based on the gathered information, selects the minimal subset of forwarding neighbors, which cover all nodes within two-hops. Each sender piggybacks its chosen forwarding nodes (i.e. multipoint relay nodes) on the outgoing broadcast packet.

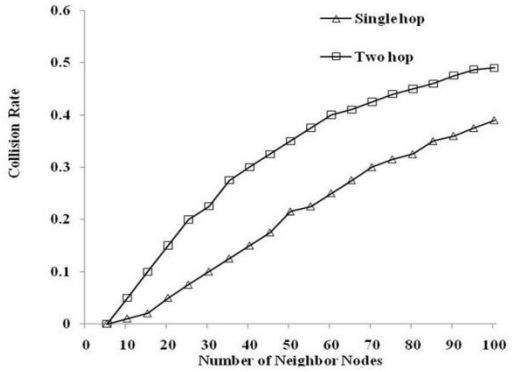


Figure 1. The collision rate of broadcast

Along the same lines, several other schemes have proposed the selection of a dominant set based on topology [18] [21] [22] [23]. All of these schemes, however, again depend on periodic hello messages to collect topological information.

The extra hello messages, however, consume resources and drop the network throughput in MANETs [14]. The extra traffic brings about congestion and collision as geographic density increases [1]. Fig 1[26] depicts the collision probability of hello messages in a single hop and a two hop network as the number of neighbor's increases. This result clearly shows that the neighbor degree causes the broadcast collision probability to increase (note, the collision probability is more than 0.1 with more than 15 neighbors). Moreover, the hidden terminal and exposed terminal problems aggravate collisions in the two hop network. Note that Fig 1 assumes no data traffic - only hello messages.

With user-data packets, the collision probability of hello messages will dramatically increase. Thus, it will be hard to collect complete neighbor topology information using hello messages. As a consequence, the aforementioned schemes (e.g., neighbor coverage, MPR, etc.) are not scalable to offered load and number of neighbors.

Lastly, we consider clustering. Clustering can be described as grouping of nodes. A representative of each group (cluster) is dynamically elected to the role of cluster head based on some criterion (e.g., lowest ID). Nodes within one hop of a cluster head become associated to its cluster. A node belonging to two or more clusters at the same time is called a gateway. Other

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

members are called ordinary nodes. Various distributed

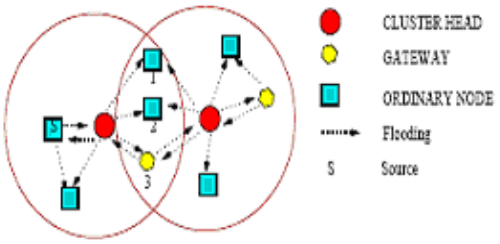


Figure 2. . An example of efficient flooding with clustering. Only cluster heads and gateways rebroadcast

computation techniques can be used to dynamically create clusters. In the active clustering lowest ID technique [15] each node attempts to become cluster head by broadcasting its ID to neighbors. It will give up only if it hears from a lower ID neighbor. Based on the above definition, any two nodes in a cluster are at most 2 hops away [9]. With this clustering scheme, the dominant forwarding nodes are the cluster heads and the gateways, as shown in Figure 2.

Clustering in ad hoc networks has been extensively studied for hierarchical routing schemes [9] [5] [3], the master election algorithms [4], power control [3], reliable broadcast [20], load aware schemes [17], efficient broadcast [16] and efficient flooding [10]. Some clustering schemes are based on the complete knowledge of neighbors. However, the complete knowledge of neighbor information in adhoc networks is hard to collect and introduces substantial control overhead caused by periodic exchange of hello messages. Passive clustering [10] is an "on demand" protocol. It constructs and maintains the cluster architecture only when there are on-going data packets that piggyback "cluster related information". Each node collects neighbor information through promiscuous packet receptions. Passive clustering, therefore, eliminates setup latency and major control overhead of clustering protocols.

Passive clustering has two innovative mechanisms for the cluster formation: First Declaration Wins rule and Gateway Selection Heuristic. With the First Declaration Wins rule, a node that first claims to be a cluster head "rules" the rest of nodes in its clustered area (radio coverage). There is no waiting period (to make sure all the neighbors have been checked) unlike for all the weight-driven clustering mechanisms [3] [5]. Also, the Gateway Selection Heuristic provides a procedure to elect the minimal number of gateways (including distributed gateways) required to maintain the connectivity in a distributed manner.

Passive clustering maintains clusters using implicit timeout. A node assumes that the nodes it had previously heard from as dead or out of its locality if they have not sent any data within timeout duration. With a reasonable offered load, a node can easily keep track of dynamic topology changes by virtue of this timeout.

Lastly, the existing passive clustering scheme uses flooding initially till the cluster is formed and each data packet carries cluster formation overhead fields even though the cluster has not changed. These limitations motivated our investigation of a new cluster formation protocol called enhanced passive clustering. While retaining the advantages of passive clustering, our scheme eliminates much of the control overhead.

III. ENHANCED PASSIVE CLUSTERING

A. *Overveiw of Enhanced passive clustering*

The proposed protocol called enhanced passive clustering is very much similar to passive clustering algorithm except that it uses RREQ packets for formation of clusters and their maintenance. RREQ packets are generated only when the topology of the network changes and hence is more suitable for cluster formation and maintenance in a rapidly changing network. RREQ packet is transmitted by the node having data to transmit but has no path to forward it. If data packets are used for cluster formation and maintenance then every data packet carries the overhead of cluster related information thus increasing the overhead transmitted in the network. Our scheme helps in reducing flooding and redundant overhead transmission effectively.

B. *Construction and maintainance*

A node can be in different states during the clustering process, namely: INITIAL, CLUSTER HEAD (CH), ORDINARY NODE, GATEWAY (GW), CH READY, GW READY and DISTRIBUTED GW.

When a node joins the network, it sets its cluster state to INITIAL. Moreover, the state of a floating node (a node does not belong to a cluster yet) also is set to INITIAL. Because enhanced passive clustering exploits RREQ packets, the implementation of this clustering resides between layer 3 and 4. An additional field in the header (the cluster information field) is carried by RREQ packet. This field contains the following entries:

- Node ID: The IP address of the sender node. Not to be confused with the source address of the IP packet.
- State of node: The cluster state of the sender node.
- Two cluster heads addresses: If a sender node is a gateway, then there is another field with the two IP addresses of the cluster heads (CHs) which are reachable from the gateway.

Below, we provide a summary description of the enhanced passive clustering algorithm.

1) The packet handling:

Upon sending a RREQ packet, each node piggybacks cluster related information in the cluster info field. Upon a promiscuous packet reception, each node extracts cluster-related information of neighbors and updates neighbor information table.

2) A cluster head declaration:

A node in INITIAL state changes its state to CH READY (a candidate cluster head) when a RREQ packet arrives from another node that is not a cluster head. With the next outgoing RREQ packet, a CH READY node can declare itself as a CLUSTER HEAD.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

3) Becoming a member (Gateway or Ordinary node):

A node becomes a member of a cluster once it has heard or overheard a message from any cluster head. A member node will serve as a gateway or an ordinary node depending on the information collected from neighbors. Specifically, a member node settles as an ORDINARY NODE only after it has learned (i.e., has heard from) enough neighbor gateways. In enhanced passive clustering, the existence of a gateway can be found only through overhearing a packet from that gateway. Thus, we define another internal state, GW READY, for a candidate gateway node that has not yet discovered enough neighbor gateways. Recall that we develop a gateway selection mechanism to reduce the total gateways in the network. A candidate gateway finalizes its role as a GATEWAY upon sending a packet (announcing its gateway's role). Note that a candidate gateway node can be downgraded to ordinary node at any time after detection of enough gateways.

C. Gateway Selection Heuristic

A gateway is a bridge node that connects two adjacent clusters. Thus, a node that belongs to two or more clusters at the same time is eligible to be a gateway. One can easily see that only one gateway is needed for the each pair of adjacent clusters. Following this observation, we have used a gateway selection mechanism that eventually allows only one gateway for each pair of neighboring cluster heads. However, it is possible that there is no potential gateway between two communicating clusters. For instance, suppose that two cluster heads are mutually reachable not by a two-hop but a three-hop route. Then the clustering scheme should select the two intermediate nodes as distributed gateways (DISTR GW).

The gateway selection mechanism can be summarized as follows:

1) Gateway

A node that belongs to two or more clusters at the same time is a candidate gateway. Upon sending a RREQ packet, a potential gateway selects two cluster heads among the known cluster heads. This node will serve as an intermediate node between those chosen cluster heads. It cannot be the intermediate gateways for any two other cluster heads that have already been announced by another neighbor gateway node. If the node finds a unique pair of cluster heads, then it finalizes its role as a gateway and announces the pair of cluster heads to neighbors.

If a gateway has received a packet from another gateway which has announced the same pair of CHs, then this node compares the node ID of itself with that of the sender. If this node has the lower ID, it keeps its role as the gateway. Otherwise, it selects another pair of CHs (that it has heard from) or changes its state to ordinary node.

2) Distributed gateway

Enhanced Passive clustering uses distributed gateway to provide connectivity among cluster heads which are 3 hops away. Moreover, distributed gateways are common at the boundary of the cluster structure. A node that belongs to only one cluster can be an ordinary node only if at least two (distributed) gateways are known to this node. Otherwise, it keeps the candidate gateway state. A candidate gateway node becomes a distributed gateway if it has not heard from any other neighboring distributed gateway belonging to the same cluster. If an ordinary node has received a packet from a distributed gateway and no other gateway is a neighbor node of that node, then this node changes to a distributed gateway.

Enhanced passive clustering was first implemented using one cluster head per cluster in DSR. In this paper this implementation is referred to as Enhanced Dynamic Source Routing protocol1 (EDSR1). Enhanced passive clustering was then implemented using two cluster heads per cluster. In this paper this implementation is referred to as Enhanced Dynamic Source Routing protocol2 (EDSR2). Two cluster heads are named as primary and secondary cluster heads. Primary cluster heads are active as long as they are in the cluster. Once the primary cluster head moves out of the cluster, instead of performing reclustering and flooding the network, this technique makes the secondary cluster head active and promotes it as primary cluster head so that flooding in the network can be avoided. It is now the responsibility of primary cluster head to select secondary cluster head. We have used lowest id technique to select secondary cluster head.

IV. SIMULATION RESULTS

The simulations were performed using Glomosim [11]. The mobility scenarios were randomly generated using random waypoint model. We used distributed coordination function (DCF) of IEEE 802.11 for wireless LANs as the MAC layer protocol.

In our simulation, 100 nodes move in a 1000x1000 meter rectangular region for 600 seconds simulation time. Initial locations of the nodes are obtained using a uniform distribution. We assume that each node moves independently with the same average speed. With the random waypoint mobility model, a node randomly selects a destination from the physical terrain and moves in the direction of the destination with a uniform speed chosen between the minimal and maximal speed. After it reaches its destination, the node stays there for a pause time and then moves again. In our simulation, the minimal speed was 0m/s and maximal speed was 16m/s and pause time was varied from 0 to 10seconds. The simulated traffic was constant bit rate (CBR).

We have analyzed the variation in the number of control packets by varying the density of the network at different pause time (i.e. mobility) of 0s, 5s and 10s.

The simulation results shown below compares the performance of standard DSR, DSR with passive clustering (PDSR) and DSR with enhanced passive clustering EDSR1 and EDSR2.

The graphs in Fig 3, 4 and 5 show that in EDSR1, the number of control packets used in routing reduces as the number of nodes increases at a fixed mobility, compared to DSR and PDSR. This is because we have used RREQ packets for clustering which limits flooding within the cluster. The number of control packets drastically reduces in the middle region of the graph as more number of RREQ packets is generated when the number of nodes increases and hence clusters are maintained efficiently. But there is slight increase in the control packets with increased mobility. This is because

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

of reclustering which causes more control packets to be flooded in the network. But the graphs for EDSR2 shows reduced number of control packets irrespective of the density and mobility of the network. As expected EDSR2 shows reduced number of control packets compared to EDSR1 even when the mobility of the node is high as reclustering is avoided. Thus we can say that using two cluster heads in passive clustering we can reduce flooding in DSR to a great extent

We have also analyzed the variation in the number of route request packets by varying the density of the network at different pause time (i.e. mobility) of 0s, 5s and 10s.

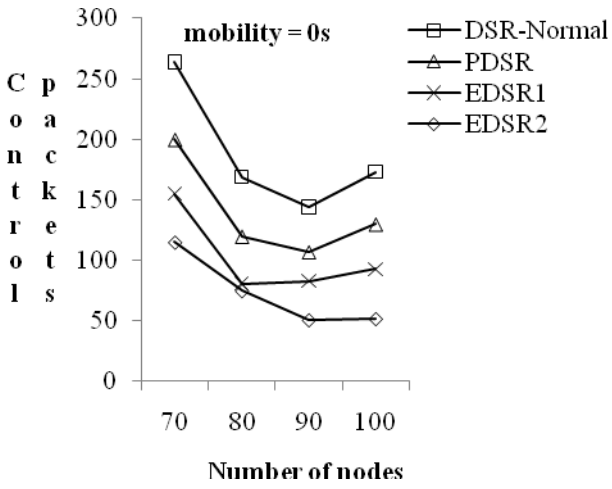


Figure 3. Number of nodes v/s control packets

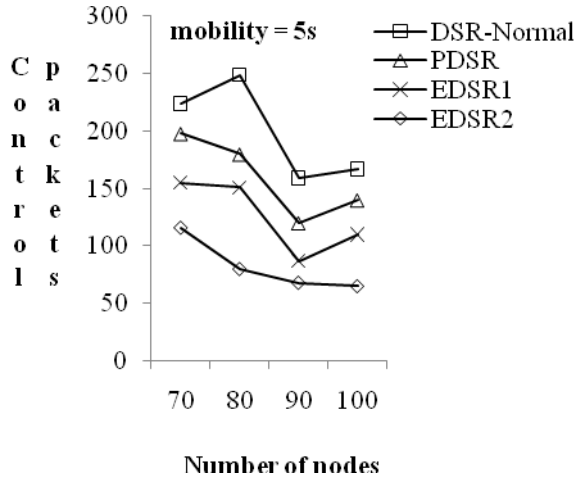


Figure 4. Number of nodes v/s control packets

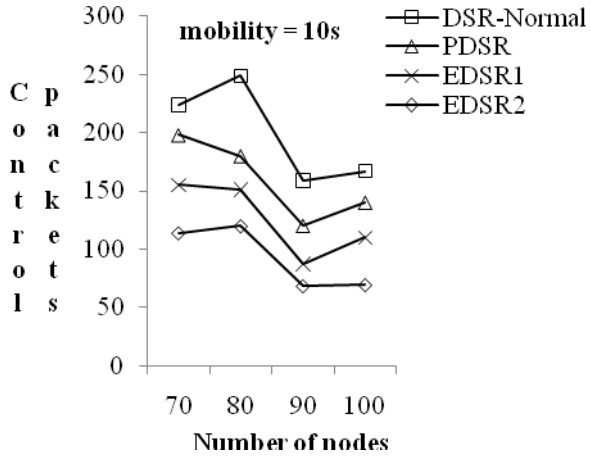


Figure 5. Number of nodes v/s control packets

The graphs in Fig 6, 7, and 8 show that the number of route requests required in the enhanced passive clustering technique is lesser than standard DSR and PDSR. This is mainly because this technique chooses the sub-optimal dominant forwarding nodes. In normal DSR we can see from the graph that there is a slight increase in the number of RREQ packets as mobility increases, this is because new routes have to be found as the position of the nodes change. Whereas in EDSR1 the graph shows that the number of RREQ packets is less for different mobilities as the routes need to be found only between cluster heads and flooding is limited to that of cluster. But in EDSR2 the number of route request packets reduces further as the secondary cluster head replaces the primary cluster head and route requests are not required to find the route between secondary cluster head as the routes will already be known during the previous transmissions.

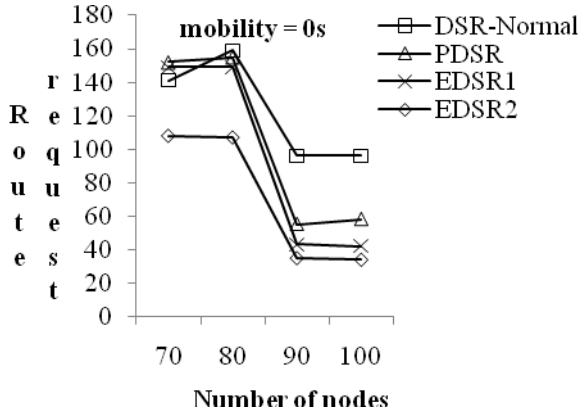


Figure 6. Number of nodes v/s Route request packets

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

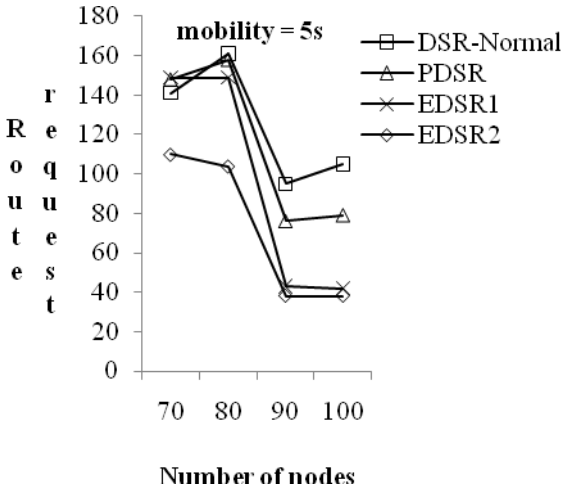


Figure 7. Number of nodes v/s Route request packets

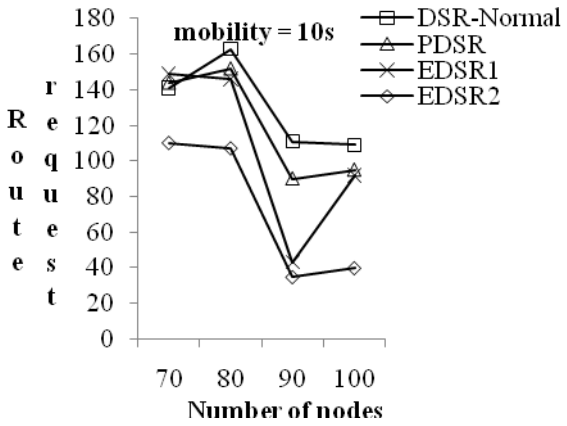


Figure 8. Number of nodes v/s Route request packets

As shown in the above graphs, increase in mobility of the nodes increases flooding in the network. This is due to reclustering. We have tried to overcome this drawback using two cluster heads.

V. CONCLUSION

For carrying out this work we have investigated the problem of "efficient flooding" based on topological information. To collect neighbor topology, the network incurs a heavy control overhead penalty - it is very costly to collect accurate topology information with node mobility and dynamically changing resources. The aforementioned topology-based schemes, in consequence, are limiting in scalability and performance due to the burden of message and processing overhead. Flooding scheme based on passive clustering removes such limitations. Passive clustering finds its home in the domain of on-demand protocols. As such protocols come alive only when the user has data and tries to find a path to send his data, and moreover they almost universally require some form of flooding; they are the ideal candidates for passive clustering. Clearly, no proactive flood enhancement scheme would make sense in this environment, as it would introduce undesirable, periodic background traffic.

Existing passive clustering scheme uses data packets for cluster formation and hence introduces lot of overhead in data transmission. So the key issue here was to evaluate the improvement introduced by enhanced passive clustering versus the original protocol version. We have applied enhanced passive clustering to the most popular reactive routing protocol DSR. Enhanced passive clustering consistently reduces the flooding overhead and improves the performance and scalability as shown by the results. This technique can also be applied to AODV, ODMRP etc.

REFERENCES

[1] Sze-Yao Ni, Yu-Chee Tseng, Yuh-Shyan C hen, Jang-Ping Sheu, "The broadcast storm problem in a mobile ad hoc network, " MobiCom, 1999.

[2] Jinyang Li, Charles Blake, Douglas S.J., De Couto, Hu Imm Lee, Robert Morris, " Capacity of Ad Hoc Wireless Networks, " Mobicom, 2001.

[3] C.R. Lin and M. Gerla, "Adaptive Clustering for Mobile Wireless Networks, " IEEE Journal on Selected Areas in Communications, Vol. 15, No. 7, Sep. 1997, pp. 1265-1275.

[4] Dennis J. Baker and Anthony Ephremides, "The Architectural organization of a mobile radio network via a distributed algorithm, " IEEE Transaction on communications, Vol. com-29, No. 11, Nov. 1981

[5] Gerla, M and Tasi, J., "A Multicluster, mobile, multimedia radio network, " ACM-Baltzer Journal ofWireless Networks, Vol. 1 No. 3, 1995

[6] Yu-Chee Tseng, Sze-Yao Ni, En-Yu Shih, " Adaptive Approaches to Relieving Broadcast Storms in a Wireless Multihop Mobile Ad Hoc Network, " Infocom, 2001.

[7] Amir Qayyum, Laurent Viennot, Anis Laouiti, " Multipoint relaying: An efficient technique for flooding in mobile wireless networks, " INRIA report, March 2000.

[8] H.Lim and C. Kim, " Flooding in wireless ad hoc networks, " IEEE computer communications 2000.

[9] Krishna, P., Vaidya, N.H., Chatterjee, M., Pradhan, D.K., "A clusterbased approach for routing in dynamic networks, " Computer Communication Review, vol.27, (no.2), ACM, April 1997.

[10] Yunjung Yi, Mario Gerla and Taek-Jin Kwon , "Efficient flooding in adhoc networks using on-demand(passive) cluster formation, " university of California. Internet Draft, draft-ietf-yi-manet-pac-00.txt, Nov. 2001.

[11] UCLA, Glomosim: A scalable simulation environment for Wireless and Wired Network systems, http://pcl.cs.ucla.edu/projects/glomosim.

[12] D.B.Johnson and D.A. Maltz, Dynamic source routing in ad hoc wireless networks, Mobile Computing, Academic Publishers, 1996.pp. 153-181.

[13] Z.J.Haas, "A new routing protocol for the reconfigurable wireless networks, "ICUPC '97 pp.562-566

[14] Giuseppe Bianchi, " Performance analysis of the IEEE 802.11 distributed coordination function, " IEEE Journal on selected areas in communications, vol. 18. March 2000

[15] R. Bagrodia, R. Meyer, M. Takai, Y. Chen, X. Zeng, J. Martin, B. Park, H. Song, " Parsec: A parallel simulation environment for complex systems, " Computer, Vol. 31(10), October 1998, pp. 77-85.

[16] M. Jiang, J. Li, Y. C. Tay, " Cluster based routing protocol (CBR) functional specification, IETF draft 1998.

[17] Yunjung Yi, Taek Jin Kwon, Mario Gerla, "A load aware routing (LWR) based on local information, " PIMRC 2001 accepted.

[18] J. Broch, D. Maltz, D. Johnson, Y. Hu, J. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network routing protocols"

[19] A.V. Aho, J.E. Hopcroft, J.D. Ullman, The design and analysis of computer algorithms, Addison-Wesley Publication. 1974.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[20] E. Pagani, G.P. Rossi, " Reliable broadcast in mobile multihop packet networks, " Mobicom 97 (1997)

[21] J. Wu and H. Li, "On calculating connected dominating set for efficient routing in ad hoc networks, " DIAL M. 1999

[22] B. Chen, K. H. Jamieson, and R. Morris, " An energy-efficient coordination algorithm for topology maintenance in Ad Hoc wireless networks, "Mobicom, 2001

[23] Y. Xu, J. Heidemann and D. Estrin,"Geography-informed Energy Conservation for Ad Hoc Routing, " Mobicom, 2001

[24] M. Seddigh, J. Solano and I. Stojmenovic, "Internal nodes based broadcasting algorithms in wireless networks, " Proceedings of the 34th Annual HICSS 2001.

[25] S.-J. Lee,W. Su, and M. Gerla On-Demand Multicast Routing Protocol (ODMRP) for Ad Hoc Networks, Internet Draft, draft-ietf-manetodmrp-02.txt, Jan. 2000.

[26] T. Clausen, P. Jacquet, A. Laouiti, P. Minet, P. Muhlethaler, A.Qayyum, L. Viennot Optimized Link State Routing Protocol, Internet Draft, draft-ietf-manet-olsr-05.txt, Nov. 2001.

[27] R. G. Ogier, F. L. Templin, B. Bellur, M. G. Lewis Topology Broadcast Based on Reverse-Path Forwarding (TBRPF), Internet Draft, draft-ietf-manet-tbrpf-03.txt, Nov. 2001.

AUTHORS PROFILE

Shobha.K.R received M.E degree in Digital communication from Bangalore University and is currently working towards the Ph.D. She is currently working as Assistant Professor with the department of Telecommunication Engineering, M.S.Ramaiah Institute of Technology, Bangalore,India. She had served as a teaching faculty at BMSCE, Bangalore for 5 years up to 1999. Her research areas include Routing protocols in Mobile Adhoc Networks and Wireless Networks.

Dr.K.Rajanikanth received M.E in Automation and Ph. D degrees from Indian Institute of Science, Bangalore, India. Areas of interest are software engineering, Object Technology, Embedded Systems. Currently working as Professor and Principal at M S Ramaiah Institute of Technology, Bangalore.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Quality of Service Enhanced Routing in Mobile Ad Hoc Networks

D.S. Thenmozhi [1], R. Lakshmipathi [2]

[1] *Lecturer, Institute of Road and Transport Technology, Erode, Tamilnadu, India.*
vivek_thenmozhi@rediffmail.com

[2] *Professor, St. Peter's Engineering College, Chennai*
Tamilnadu,India.

***Abstract*—The importance of Mobile Ad hoc Networks (MANETs) has been increasing everyday. Attractive features of MANETs are no need for infrastructure and decentralized nature. Many applications that use MANETs include multimedia data that require Quality of Service (QoS) support for effective communication. Also QoS routing feature is important for a mobile network to interconnect wired network having QoS support. We approach the problem of providing Quality of Service for mobile ad hoc networks (MANETs) by the technique of bandwidth based path finding. The Ad hoc On demand Distance Vector routing protocol provides efficient route establishment between nodes with minimal control overhead and minimal route acquisition latency. The normal route finding method of AODV is improved as Quality of Service Enhanced Ad hoc On demand Distance Vector (QEAODV) routing. In our work QEAODV establishes a path between the source and the destination meeting the application stipulated throughput requirement. Contention which is the inherent problem in MANET is considered effectively in QEAODV. QEAODV is implemented so that additional overhead requirement will be very less. In this paper, we present a scalable and efficient QEAODV to support QoS in ad hoc networks. Simulation results show significant performance advantages of our protocol when compared with normal AODV.**

***Keywords*—Mobile Ad hoc Networks (MANETs), quality of service routing, bandwidth estimation, contention, admission control, simulations.**

I. INTRODUCTION

Ad hoc networks are communication networks formed by a number of nodes which are small radio devices with limited computational capacity and memory. The most desirable advantage of ad hoc networks are their easy deployment. Ideally it should be possible to deploy the nodes in the area of operation and have them self-organize to route traffic as required. Such an easy deployment would be most advantageous in a variety of applications ranging from military operations and disaster relief to commercial applications.

MANETs (Mobile Ad hoc Networks) [12] inherently possess many challenges for its easy deployment, the nodes should not depend on an external energy supply, so they are normally battery powered, and battery life is often a limiting factor. The radio transmission channel is limited in bandwidth. Channel bandwidth is shared among neighboring nodes. Nodes in the MANETs can move freely. Nodes mobility makes determining and maintaining the network topology the most challenging issue. Networking mechanisms such as routing protocols for MANETs require high efficiency because of limited resources in a mobile node such as network bandwidth, memory capacity and battery power. Routing protocols in ad hoc networks must manage frequent topology changes caused by node mobility. The nature of dynamic changing topology in ad hoc networks introduces difficulties in end-to-end route finding. A new technique designed for ad hoc networks is 'On-demand', or reactive routing. In this technique nodes are not maintaining the full topological view, but nodes construct routing tables having only routes to nodes that a source needs to communicate with, that are established on demand with the help of source flooding.

A lot of work has been made on routing in ad hoc networks: the destination - sequenced distance vector (DSDV) protocol [5], the wireless routing protocol [6], the temporally-ordered routing algorithms [7], the dynamic source routing protocols [8], the associativity based routing protocol [9], and the zone routing protocol (ZRP) [10], etc. These protocols tend to establish a path with least number of hops and achieving a high degree of availability of nodes involved in the active path where the network topology changes quickly. Also, all the previous routing solutions only deal with the best-effort data traffic.

Though, the vast array of technological solutions for ad hoc networks, their practical implementation and use in the real world has been limited so far. Because entertainment and other multimedia based applications are naturally what drive the mass users of a technology. In order to support above services best-effort routing solutions are not sufficient. Normally multimedia applications often have stringent delay and reliability sensitive service requirements. Any networks supporting multimedia applications must cater above requirements. Hence focus has been shifted from best-effort services to the provision of better defined QoS in ad hoc networks. Shared nature of the medium in MANETs needs additional care at the time of Quality of Service support. QoS supportive routing protocols [2], [3] find important role in

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

QoS mechanisms. Since their role is to find which nodes in the network will meet application requirements in the wireless, mobile ad hoc network environment.

II. Related Works

In MANETs Quality of Service based routing is a relatively new problem. In this section we present an overview of the existing solutions. Chen and Nahrstedt [4] proposed a ticket based QoS routing algorithm for ad hoc networks. This ticket based probing scheme achieves a balance between the single-path routing algorithms and the flooding algorithms. It does multipath routing without flooding. The required QoS is ensured during the time when an established path remains unbroken. The QoS support however is disrupted during the rerouting time. They consider only the type of networks whose topologies are relatively stable because their routing, rerouting architectures do not support ad hoc networks with violently changing topologies. Lin and Liu [11] proposed a new bandwidth routing scheme which contains bandwidth calculation and reservation for mobile ad hoc networks. They suggested a TDMA-based approach. This approach requires effective synchronization between all nodes in the networks. Applying highly synchronized solutions in ad hoc networks becomes expensive and synchronization may fail when the nodes are mobile. The free slot allocation algorithm based on TDMA scheme is vulnerable to node mobility in the networks since slot allocations must be reconfigured whenever there are changes in available bandwidth or changes to routes in the network. Also they have not considered the routing optimality (i.e. shortest path). Hanzo et. al [14] proposed Quality of Service routing in ad hoc networks. They suggested throughput constrained Quality of Service routing utilising the Dynamic Source Routing (DSR) protocol. DSR is based on source routing in which each data packet carries complete path information. This requires more overhead compared to AODV.

Normally QoS can be achieved by coordinating the transmission schedule of packets between nodes. Actually existing above discussed approaches are mostly based on local decisions. All are focusing on the packet level and only deal with required resource allocation at individual node point of view. In order to support Quality of Service, guarantees for end to end flows, path finding approaches need to be combined with suitable admission control strategy. At the time of making admission control decisions, a node considers its local resources simultaneously it must account the resource of this contention neighbors because nodes flow may consume their resources through contention. This paper fulfils this objective by modifying AODV to perform admission control logic at every node and also to consider both node's local resources and resources at contention-neighbors while making admission control decisions. Admission control is based on the knowledge of available bandwidth information. We only consider bandwidth as the admission criteria. This is because bandwidth guarantee is one of the most critical requirement for real time applications. The performance of Quality of Service Enhanced AODV is compared with its counter part AODV using NS-2 simulator.

The rest of the paper is organized as follows. Next section presents basics of QEAODV protocol. Section IV explains the complete functioning of Quality of Service Enhanced AODV. Section V deals with QEAODV route discovery process. The protocol simulation and results are discussed in section VI.

III. Protocol Description

Basic AODV [13] is based on flooding the network with Route Request (RREQ) messages. The source node broadcasts a RREQ message with a time-to-live value equal to 1. i.e. a broadcast is limited one hop neighborhood. Each RREQ is uniquely identified through a sequence number, so that the first copy of a RREQ received by a node is processed, while duplicated messages are discarded. When a node receives the first copy of a given RREQ it records the address of the node that sent the message. When the first RREQ reaches the desired destination, a route reply (RREP) message is generated and sent back to the source node through the recorded reverse path, ensuring a path from the source to the destination. Normally this approach minimizes the number of hops of the chosen path. The basic functionality of the QEAODV is much similar to the AODV protocol. QEAODV differs from AODV in the way the route discovery process is modified to provide quality of service support by performing bandwidth constrained admission control at each node in the network. Similar to AODV, the QEAODV also uses the Route Request, Route Reply and Route Error packets for the route discovery and maintenance process, except the Route Request and Reply packet formats are modified to carry additional information through the network

IV. Functioning of QEAODV

The main problem of the MANET comes from the shared nature of the wireless medium in single-channel networks. Essentially, nodes that cannot communicate with each other directly may still contend directly with each other for the same resources. This extended contention area, known as 'neighborhood contention' affects resource allocation at individual nodes in two-ways. First allocation decisions at an individual node require bandwidth information of nodes outside of its communication range and along the entire route. Second, contention for resource may involve multiple nodes along a route. QEAODV performs admission control based on knowledge of these characteristics of MANET. We focus on ad hoc networks based on single-channel MAC layers like IEEE 802.11 because these single channel protocols are widely available and typically support ad hoc communication. Moreover, these protocols are simple to implement and robust and do not rely on stringent time synchronization that is hard to implement in ad hoc network. The physical characteristics of wireless channels introduce the two challenges. First

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

challenge is available bandwidth estimation at a node, second challenge is estimation of flow bandwidth requirement in a shared medium.

A. Node's Available Bandwidth:

In shared wireless medium, when a node starts to transmit a flow, it consumes bandwidth from its contention neighbors. Because each node has a different view of the network, the node cannot decide on its own whether its contention neighbors have sufficient unused bandwidth for the new flow. Also, obtaining contention neighbor information is not easy since a node may consume the bandwidth of contention neighbor but not able to directly communicate with that neighbors if that neighbors are located outside transmission range and inside carrier-sensing range.

B. Flow Bandwidth Consumption:

Multiple nodes on a route may contend for bandwidth at a single location and not know about each other. A node on the route of flow cannot tell how much bandwidth the flow will consume at its contention neighbors. The limited bandwidth of wireless ad hoc networks requires the limitation of any message overhead from information collection. In addition, due to mobility, information gathered about the network only has limited lifetime. Hence, it is effective to collect information as close as possible to the time and location that it is needed.

C. QEAODV Admission Control:

The objective of admission control is to determine whether the available resources can meet the requirements of a new flow while maintaining bandwidth levels for existing flows. Each node views a different channel state, the available bandwidth in the network is not a local concept. To tackle this condition, two terms are introduced: local bandwidth available (BWlocal), contention-neighborhood bandwidth available (BWc-neigh). Local bandwidth available is the amount of unconsumed bandwidth as observed by a given node. Contention neighborhood available bandwidth is the maximum amount of bandwidth that a node can use for transmission without affecting the reserved bandwidth of any existing flows in its carrier-sensing range. Since a node may consume the bandwidth of nodes that are with in its contention range, the contention neighborhood bandwidth available for a given node is equal to the smallest local available bandwidth of all its contention neighbors. Hence in order to admit a flow, a node must have required local bandwidth and contention neighborhood bandwidth.

1. Calculation of Local Bandwidth Available (BWlocal)

It is the unconsumed bandwidth at a given node. Each node in the MANET can determine its BW_{local} by passively listening network activities. In our approach, we use the fraction of channel idle time based on the past history as an indication of local available bandwidth at a node. This approach is justifiable since it does not consider that some of the channel time cannot be used due to the idle time caused by the exponential backoff algorithm in IEEE 802.11 and the collisions in the network. Using the fraction of idle channel time can be a simple approximation for local available bandwidth. A node can perceive the channel as either idle or busy. The channel is idle if the node is not in any of the following three states. First, the node is transmitting or receiving a packet. Second, the node receives a RTS or CTS message from another node, which receives channel for a period of time specified in the message. Third, the node senses a busy carrier with signal strength larger than a certain threshold, called the carrier-sensing threshold, but the node cannot interpret the contents of the message. By monitoring the amount of channel idle time, T_{idle}, during every period of time, T_p, the local bandwidth available BW_{local} , for a node can be computed using a weighted average [1] as follows

$$BW_{local} = \omega\, BW_{local} + (1-\omega)(T_{idle}/T_p)\, BW_{channel} \quad (1)$$

Where $BW_{channel}$ is the capacity of the channel and weight $\omega \in [0,1]$.

2. Calculation of Contention Neighborhood Bandwidth Available ($BW_{c\text{-}neigh}$)

Each node perceives the network in a different state. Hence a node's local bandwidth available cannot provide information about its contention neighbors. Since it does not know the amount of BW_{local} at other nodes. Two approaches are normally used to obtain bandwidth information at contention neighboring nodes. They are active approaches and passive approaches. In active approaches, neighbors voluntarily exchange bandwidth information between each other. Such exchanges normally incur high message overhead. In passive

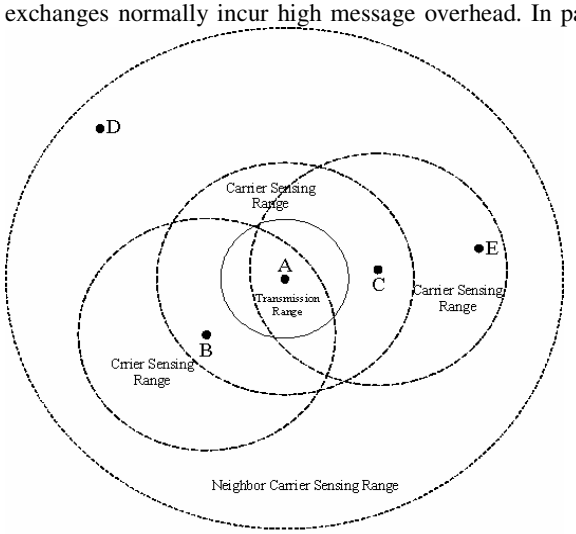


Fig. 1. Various sensing ranges of a mobile node

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

approaches, a node passively monitors the channel to calculate its contention neighbors' local bandwidth available. During the normal IEEE 802.11 operations, a node listens the medium using a threshold value known as contention-carrier sensing threshold, which is set to a value much lower than the carrier-sensing threshold refers the range that covers the carrier-sensing ranges of all of the sensing node's contention neighbors as shown in the Fig.1. When the signal strength of the carrier sensed by a node is smaller than the contention carrier sensing threshold there is no communication in its contention neighborhood and contention neighbors of the node experience idle channels. The amount of time that the channel is in this idle state, denoted as $T_{idle}^{contention}$, for every period of time, T_p, contention neighborhood available bandwidth, BW_{c_neigh} is calculated using the following formula.

$$BW_{c\text{-}neigh} \approx \omega\, BW_{c\text{-}neigh} + (1-\omega)(T_{idle}^{contention}/T_p)BW_{channel} \quad (2)$$

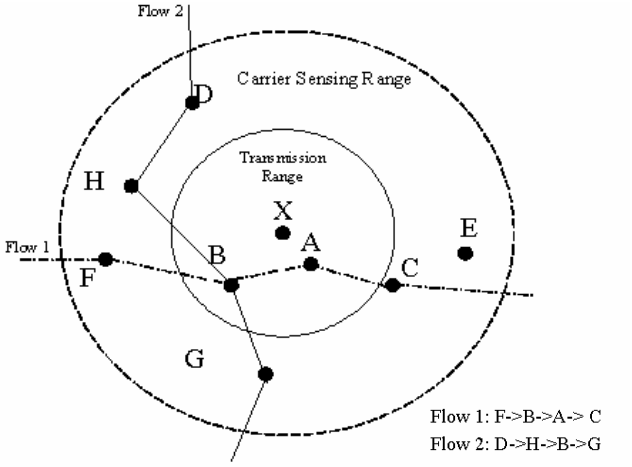


Fig. 2. Channel contention at a node in a multihop flow.

3.Calculation of Application's Flow Bandwidth Consumption (BW_{a_flow})

QEAODV needs to quantify the bandwidth that a new flow requires so that it can be decided whether the bandwidth available will satisfy the requirements of the flow. Foremost, the application's data rate has to be converted into the corresponding channel bandwidth requirement. This conversion includes the protocol overhead incurred in the MAC layer and the network layer. As per IEEE 802.11, for every application data packet, the MAC layer performs handshaking. During this RTS, CTS and ACK control packets are involved. Hence each data packet's transmission time is calculated as follows.

$$T_{data} = T_{rts}+T_{cts}+T_{ack}+T_{difs}+3T_{sifs}+(P+Q)/BW_{channel} \quad (3)$$

Where T_{data} - transmission time of each data packet
T_{rts} - time for transmitting RTS
T_{cts} - time for transmitting CTS
T_{ack} - time for transmitting ACK
T_{difs} - DCF inter frame space defined in the IEEE 802.11 protocol standard
T_{sifs} - short inter frame space defined in the IEEE 802.11 protocol standard
P - size of the data packet
Q - IP and MAC packet header length
$BW_{channel}$ - Channel capacity

If at every second, the application generates 'R' packets with average packet size 'P', the corresponding channel bandwidth requirement is computed as follows.

$$BW_{flow} = R \times T_{data} \times BW_{channel} \quad (4)$$

Next important factor to be considered is multiple nodes on the route of a new flow may contend for bandwidth at a single location. Every such node needs bandwidth equal to BW_{flow}. The number of such kind of nodes is known as contention count (C_{ct}). Hence the bandwidth consumption of the flow [1] at this location is expressed as:

$$BW_{a\_flow} = C_{ct} \times BW_{flow} \quad (5)$$

In Fig.2, flow 1 passes through the path F->B->A->C. Flow 2 passes through the path D->H->B->G. Both the flows are within node X's contention range. Assume flow 1 requires 1 Mbps at each node, flow 2 requires 0.5 Mbps at each node. At node X, both flows will take off 6 Mbps because nodes involved in both the flows are within the carrier sensing range of node X. A node can learn its contention neighbors by passively listening the routes and information carried in control and data messages. This method does not impose any extra message overhead in the network. A node on the route can get its contention count (C_{ct}) value if the complete path information is available to it. Complete path information is made available in the route reply (RREP) packet of QEAODV. Hence BW_{a_flow} is computed (5) by a node on receiving RREP packet. Final admission control at a node is based on this value

V. QEAODV ROUTE DISCOVERY PROCESS

Like AODV [2], QEAODV is a reactive unicasting routing protocol for mobile ad hoc networks. It needs to maintain the routing information about the active paths. In QEAODV, routing information in maintained in routing table at every node. Each node constructs a next-hop routing table, which contains the destination to which it currently has a route. An entry in the table automatically expires, if it has not been used for a specified amount of time.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

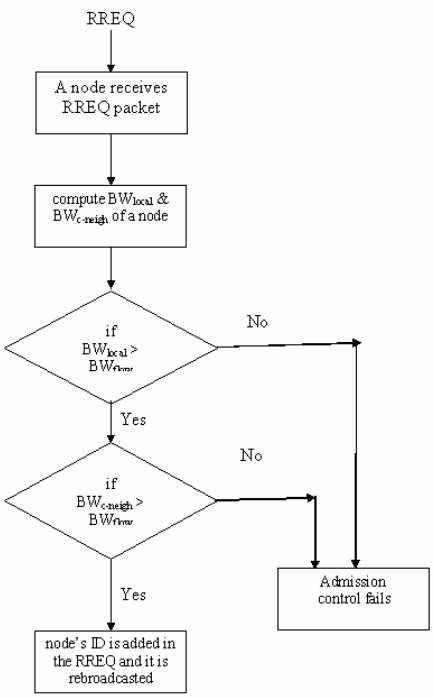


Fig. 3. Preliminary Admission Control

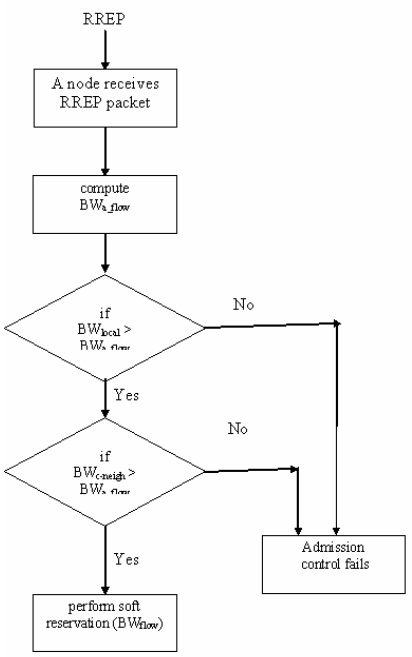


Fig. 4. Final Admission Control

During the route discovery process, the source broadcasts route request (RREQ) packets. Each RREQ packet contains the addresses of the source and the destination, the broadcast ID, the last seen sequence number of the destination as well as the source node's sequence number. Broadcast ID is used as an identifier. Sequence numbers are utilized to ensure loop-free and up-to-date routes. Application's channel bandwidth requirement (BW_{flow}) is computed by the source and included in the RREQ packet. In QEAODV, each node computes BW_{local} and $BW_{c\text{-}neigh}$ as per (1) and (2) respectively. Every intermediate node, on receiving RREQ performs preliminary admission control as given in Fig.3. If the bandwidth requirement of the flow BW_{flow} is lower than node's local available bandwidth BW_{local} and contention neighborhood available bandwidth $BW_{c\text{-}neigh}$, preliminary admission control succeeds, otherwise it fails. In case of failure, the RREQ is discarded. On success of the preliminary admission control the node sets up a reverse route entry in its routing table, adds its identifier in the RREQ packet and rebroadcasts the route request. Recording the sequence of hops in RREQ packet enables to determine the lower bound of the contention count of the complete route and also it can be used to eliminate circular routes.

When the intended destination receives, a route request, it receives the full route and sends a route reply (RREP) back to the source along the same route. If different routes arrive at the destination, the destination chooses the shortest path route and the remaining routes are cached at the destination for certain time period in order to make use of them if final admission control fails for the first selected route. On receiving the RREP packet, a node performs final admission control as shown in Fig.4. Using the details available in the RREP packet each node on the reverse path computes actual bandwidth consumption of the flow ($BW_{a\_flow)}$. If the actual bandwidth requirement of the flow ($BW_{a\_flow)}$ is lower than node's local available bandwidth (BW_{local}) and contention neighbourhood available bandwidth ($BW_{c\text{-}neigh}$), final admission control succeeds otherwise it fails. On success of admission control, a soft reservation of bandwidth is made in the routing table and RREP is forwarded to its immediate predecessor. On failure of final admission control, an

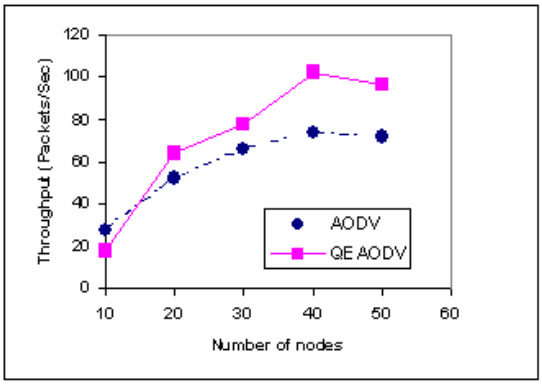


Fig. 5. Throughput of QEAODV.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

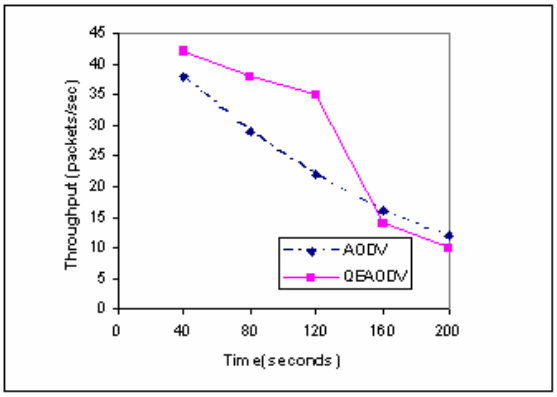


Fig. 6. Throughput for different simulation times.

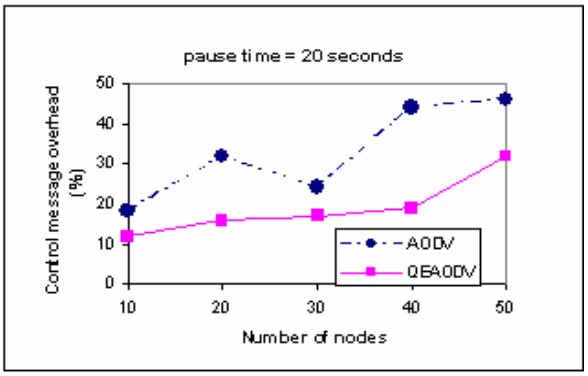


Fig. 8. Control message overhead for pause time = 20sec

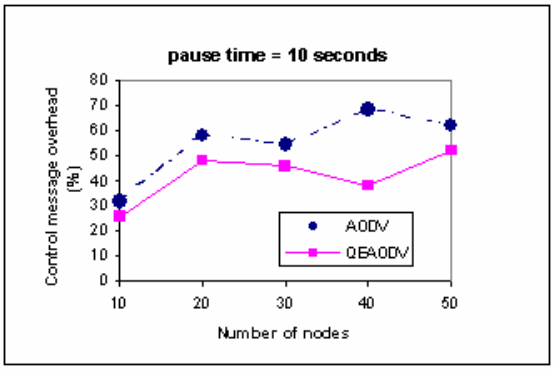


Fig. 7. Control message overhead for pause time = 10sec

admission failure message is sent to the destination via the same reverse route. It enables cancellation of bandwidth reservation by the successor nodes. On receiving the admission failure message, the destination selects another fresh cached route and sends a RREP. On successfully receiving RREP, a source has enough end-to-end bandwidth reserved entire route and communication can start at each node on the path, bandwidth reservation is refreshed by the arrival of data packets. The bandwidth reservation at the node automatically expires, if no data packet arrives due to link breakages.

VI. Simulation and Results

The proposed QEAODV routing protocol is implemented using the NS-2 network simulator [15] by modifying the code of AODV protocol. AODV protocol already exists in the network layer. A modification is done in the MAC layer to capture the signal strength. In QEAODV the packet structure of RREQ is modified to carry additional information. The routing table structure is also changed to hold the extra details.

Simulations are run for different scenarios. Different scenarios are created using 10, 20, 30, 40 and 50 nodes. Protocol evaluations are based on the simulation of wireless nodes forming an ad hoc network, moving about over a rectangle. Rectangle size is 1000m x 1000m, simulation time is 200seconds. At medium access control (MAC) layer the 802.11 protocol was used. Radio transmission range of a node is set to 250m and the carrier sensing range is set 550m. Node movement is set as per "random way point" model. Each flow generated 10 packets per second. Each packet size is 512 bytes. Speed of nodes is 5m/s and the bandwidth of the channel is 2 Mbps.A number of simultaneous CBR flows are made. Scenarios are run for different node pause time values. The performance of the QEAODV is compared with normal AODV in terms of throughput, overhead and network performance ratio. Throughput of QEAODV gets increased significantly as shown in Fig. 5. In Fig. 6 throughput of QEAODV is high till the average simulation time of 130 seconds, beyond that throughput gets decreased. This is due to the effect of nodes mobility. From Fig. 7 and Fig. 8, it is inferred that control message overhead as per QEAODV execution is comparatively lower than normal AODV execution. Also as the nodes' pause time value increases, control message overhead decreases drastically.

VII. Conclusion

In this paper, we proposed a QoS enhanced AODV (QEAODV) routing algorithm for ad hoc networks. The existing AODV performs routing with low control overhead and effective packet transmission. But do not have QoS support. We modified the normal AODV to perform path finding that meets the application stipulated bandwidth requirement. Our path finding approach is modified in such a way that it deals with common medium sharing problem of

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

the ad hoc networks effectively. The modified QEAODV performs path finding with less overhead by adopting passive approach of listening to the medium. Simulations show that QEAODV performs better than AODV in terms of throughput and control message overhead. It improves packet delivery ratio greatly without affecting the overall end-to-end throughput of existing flows.

REFERENCES

[1] Y. Yang and R. Kravets, "Contention-aware admission control for ad hoc networks," IEEE Trans. Mobile Comput., vol. 4, pp. 363-377, Aug 2005.

[2] Z. Fan, "QoS routing using lower layer information in ad hoc networks," in Proc. Personal, Indoor and Mobile Radio Communications Conf., pp. 135-139, Sep. 2004.

[3] L. Hanzo and R. Tafazolli, " A Survey of QoS Routing Solutions for Mobile Ad Hoc Networks," IEEE Communications Surveys, vol. 9, No. 2, 2nd Quarter, 2007.

[4] S. Chen and K. Nahrstedt, "Distributed quality-of-service routing in ad hoc networks," IEEE J.Select. Areas Commun., vol. 17, pp. 1488-1505, Aug. 1999.

[5] C. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance vector routing (DSDV) for mobile computers," in Proc. ACM SIGCOMM'94, pp. 234-244.

[6] S. Murthy and J. J. Garcia-Luna-Aceves, "An efficient routing protocol for wireless networks," ACM Mobile Networks Applicat. J., Special Issue on Routing in Mobile Communication Networks,1996.

[7] V. D. Park and M. S. Corson, "A Highly adaptive distributed routing algorithm for mobile wireless networks," in Proc. IEEE INFOCOM'97, pp. 1405-1413.

[8] D. Johnson and D. Maltz, "Dynamic source routing in ad hoc wirelesss networks," Mobile Computing. E. Imielinski and H. Korth, Eds. Norwell, MA. Kluwer, 1996.

[9] C. K. Toh, "Associativity-based routing for ad hoc mobile networks," Wireless Personal Commun., vol. 4, pp. 103-139,1997.

[10] Z. J. Haas and M. R. Pearlman, "The performance of query control schemes for the zone routing protocol," in Proc. IEEE Symp. Computers and Communication, 1998.

[11] C.R.Lin and J.S.Liu, "QoS routing in ad hoc wireless networks," IEEE J. Select. Areas Commun., vol. 17, pp.1426-1438, Aug. 1999.

[12] C.E.Perkins, ed., Ad Hoc Networking, ch. 3. Addison Wesley, 2001.

[13] C. E. Perkins and E. M. Royer and S. R. Das, "Ad Hoc on Demand Distance Vector (AODV) Routing," Internet Draft, draft-ietf-manet-aodv-08.txt, 2001.

[14] L. Hanzo (II) and R. Tafazolli, "Quality of Service routing and admission control for mobile ad hoc networks with a contention-based MAC layer," in Proc. 3rd IEEE Conf. Mobile Ad hoc and Sensor Systems, (Vancouver), pp. 501-504, Oct. 2006.

[15] S. McCanne, S. Floyd, S. Fall, K. Varadhan: The Network Simulator NS2 (1995). The VINT Project, available at http://www.isi.edu/nsnam/ns.

PC 2 Phone Event Announcer

BEHRANG PARHIZKAR
Faculty of Information & Communication Technology, LIMKOKWING University
Cyberjaya, Selangor, Malaysia
hani.pk@limkokwing.edu.my

ABDULBASIT MOHAMMAD ABDULRAHMAN ALAZIZI
Faculty of Information & Communication Technology, LIMKOKWING University
Cyberjaya, Selangor, Malaysia
Basit5050@yahoo.com

MOHAMMAD NABIL SADEGH ALI
Faculty of Information & Communication Technology, LIMKOKWING University
Cyberjaya, Selangor, Malaysia
Nabil1420@yahoo.com

ANAND RAMACHANDRAN
Faculty of Information & Communication Technology, LIMKOKWING University
Cyberjaya, Selangor, Malaysia
anand@limkokwing.edu.my

SUJATA NAVARATNAM
Faculty of Information & Communication Technology, LIMKOKWING University
Cyberjaya, Selangor, Malaysia
sujata@leadership.edu.my

***Abstract*-Nowadays, mobile phones are indispensable devices; it has become a trend now that college and university students are owners of such devices and this particular factor plays a very important role behind the coming up with the proposed system. "PC 2 Phone Event Announcer", is the name of the new proposed system suggested to solve the existing communication problem between the College staff and students. As the name suggests, it can be deduced that the system will involve computers and phones, more specifically mobile phones.**

Keywords-component; Mobile; SMS; Phone;

I. INTRODUCTION

Mobile phones play an essential role in the mobile communication of the modern world. They make a person easily accessible. 'PC 2 Phone Event Announcer' makes it possible to access a group of persons in just a few clicks. The system will permit lecturers, records office staff and the librarian to contact students by sending SMS through their computers. The records office and the library will have a special feature called Autorun. This feature will do the job of the specified staff members by automatically scanning through the database for any late payments or overdue books and send SMS or emails to the concerned students.

In this paper we are going to conduct a study of the current system on the basic definition and concepts about Pc 2 Phone event announcer systems. Furthermore, we have done a small investigation about this topic. In the present scenario each and every person uses mobile phones. They have become a trend, students from every college and university are the owners of mobile phones and this particular factor plays a very vital role behind this project. Mobile phones play an important role in the mobile communication of the modern world. They make a person to contact with others very quickly.

II. RELATED WORK

In this ever increasing world of communication technology, the use of computers and cell-phones is getting wider and wider day by day. From the time the cell-phones were introduced to the market, the PCs and hand-phones have been increasingly sharing the similar applications and features. Nowadays, although being dependant from each other, they're being interconnected and inter-dependant in several ways. The following literature review tries to brief on several different studies that have been done on this interconnectivity of PCs and cell phones in different respects and also on different ways that cell-phone services have facilitated people's lives.

A research is conducted to find out how SMS technology can be used for as means of quick communication in particular the bulk SMS services used in organizations. We will go through many

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

articles on SMS technology utilization by various organizations as means of communication.

EMERGENCY TEXT COMMUNICATIONS:

Being an absolutely useful communicational tool, text communications can specially be used in cases and places where there's no possibility of using voice calls. Wonsang Song, Jong Yul Kim, and Henning Schulzrinne, Piotr Boni and Michael Armstrong focused on the advantages of using text communications like IM and SMS in Emergency services. They focused on such features like being fast in service, text-based which can be accompanied with multimedia and Automatic Geographic Location in their article. Among the advantages of this service they pointed at its easy usability especially for the deaf and in places in which there's no chance of having a voice call. Also they mentioned that using this service, the current network can be maintained and it does away with the need to broaden the network bandwidth and so is cost-efficient [1].

AUTOMATED SECURED APPLIANCES:

Women University Rawalpindi of Pakistan has studied on a remote SMS-based system of controlling home appliances while the user is away from home in 2009. Using this system, users can have a global access to their home appliances and this feature enables them to automate their homes and secure them against intrusion with a relatively low cost. This is done through SMS using GSM technology. Since this system is wireless it's so adaptable and cost-efficient. Some technologies used here are GSM, RF (Radio Frequency), GPRS Modem, and Cell-phone via serial port RS 232 [2].

Also Fdhil T. Aula (2005) studied the same issue and discussed the techniques of using SMS for controlling home appliances through PC Parallel Port Interfacing [3].

SMS INTEGRATION IN ENTERPRISES:

In his article, Daniel Mavrakis, have reviewed different ways that PCs and cell-phones can be integrated into firms' routine chores using SMS. The featured mentioned by him are Alarm massages, Remote Control, Machine-To-Machine Interface, and Location Retrieval [4].

E-HEALTH ENHANCEMENT USING MOBILE COMMUNICATION:

Healthcare industry's increasing use of Mobile technology has been sparkling healthcare initiatives round the world. V. Dinusha and K.D Arunatileka focused on those aspects of mobile communication used in e-health like, sending scheduled SMS, replying and auto forwarding and RS232Serial Data cable. The technologies used here are Bluetooth, GSM, GPRS, EMR (Electronic Medical Record), infrared port, Serial cable connection, and RS232 serial data cable. Speaking about the advantages of using mobile communication in e-health, they believe that it enables, excels and aids accurate predictions, exact analysis, updates on diseases via SMS, timelier, a more comprehensive public health information and improved system of diagnosis disease tracking [5].

SMS-BASED PERSONAL ELECTROCARDIOGRAM MONITORING SYSTEM:

Ashraf A. Tahat studies the advantages of using SMS at the time when there's an emergency with Electrocardiogram Monitoring System. Features and technologies used in this article are Bluetooth Transceiver, The Microcontroller, Body Temperature Sensor, GPRS, EDGE, 3G, and WIMAX. In case of an emergency, when there's no immediate access to a cardiologist, ECG electrodes can be attached to the patient's body and the necessary information will be sent to the cardiologist's hand-phone, PDA or PC and a fast diagnosis and decision can be made by him/her [6].

SMS SERVICE FOR SCHOOLS/CONFERENCES:

James Kadirire explains in his article the ways that teacher-student relationship can be facilitated through the use of SMS, sent from PC to a phone and visa a versa. Featured used in this article are Internet, Email, Auto alerts, and Wierless and the technologies are JavaServlets, SMS Gateway, Unix Platform, and GSM Modem. The advantages of using such systems explained by Dr. Kadirire the facilitation of Presentations and Conference, the ability to maintain Instant Services, and elimination of cable usage due to the use of wireless systems [7].

SMS-BASED E-GOVERNMENT MODEL:

E-Governance has been getting a growing concern for the governments round the world especially in the developed and developing countries. It can be of great help to governments because of its potential for boosting up the official processes and doing away with much of bureaucracies. Toney Dwi and Robert proposed in their article some current technologies for the implementation of SMS-based systems in the application e-governance model. Some features used in their SMS-Based E-Government Model are Event-Based and Scheduled-Based SMS and the used technologies are SMS Cell Broadcast (CBS), SMSC and HTTP Protocols, ISP or ADSL, and Short Message Driver. Some of the advantages of this model mentioned by Susanto and Goodwin are the receipt of alerts at scheduled times (for scheduled

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

SMS), user's ability to request services by SMS, and the probability of having a Personal Profile [8]. This model maintains that local authorities can enhance their SMS-based governance through five levels: Notification, Presentation, Communication, Transaction, and Integration. Each level is a kind of e-government service that can be offered using SMS services.

MOBILE PHONES IN EDUCATION:

Michael Yerushalmy & Oshrat Benzaken has reviewed the ways that Mobile phone services can facilitate the education. Technologies used here are GPRS Modem, 3G, WiFi, J2ME, Medlets, and IMode. These technologies are used to send Instant Messages to mobile phone, and using those technologies this sending of IMs can be done with a PC using a GPRS Modem. The main advantage of using this technology is coming to the users' present location and their availability [9]. Classmates can easily get in touch with their peers and teachers, do a group discussion, ask for personalized information and teachers can check the progress of their students with no need of physical presence.

Saranphong Pramsane & Ridwan Sanjaya also believes that a PC to phone communication through SMS can be used in educational systems. The use of this system would result in closer relationships between teachers and students. They use technologies like WAP, GPRS and GSM Modem, HTTP, SMPP. The advantages of using such PC to phone messaging technologies are remote access, reliability in packet transfers, and the fact that HTTPs are acceptable in all browsers. Also they've considered the fact that not only texts but also multimedia files can be transferred from PCs to hand-phones using such technologies. This issue is of high importance and value in educational systems. [10] Tutors can send students whatever materials they need and this way of transmittance of the data is so cost-effective and efficient which any student can afford it.

Yet other researchers studied other different applications of PC-to-phone interconnections and mobile phone services. O. Awodele, E. R. Adagunodo, A.T. Akinwale, S. Idowu and M. Agbaje (2009) explained the use of cell-phone for delivering the exam results through SMS [11].

Didier Chincholle, Michael Bjorn, Christian Norlin & Morgan Lindqvist (2006) researched about the development of a user-centered IMS-based system through a user-oriented approach [12]. Rajive Chakravorty spoke about the limitations of using GSM and GPRS technologies in PC-to-phone communications and the advantages of using a TCP (Transmission Control Protocol) over GPRS [13]. In a similar study, Laurie Butgereit explained the usefulness of GPRS technology in Instant Messaging and how IMs can be sent using GPRS. [14] Neseem Alrawi worked on the use of use of Mobile SMS through GPRS and Java technologies in computer event communications [15]. Kjell suggested the use of Bluetooth and J2ME in sending and receipt of SMS and IM from PC to phone [16]. Also Klinkman studied the use of GSM Control SMS Gateway in sending message to hand-phones from PCs. [17] Finally, In an interesting study, Mahesh Gogineni & Aishwarya Lackshmi discussed the techniques and advantages of SMS in mobile phone that are capable being applied to urban microfinance [18].

III. PROPOSED SYSTEM

The proposed system will work easily; the lecturers will have a system that will **enable** them to send SMS to their students from the web. They will have a Login ID and a password allowing them to access to the system. A student database consisting of student information such as their phone numbers, ID and name will be provided in order to facilitate the job of the lecturer to choose the students. After lecturer successfully login he can send Pc 2 Phone Event Announcer information to students by SMS. They will also have access to their respective timetables that will help for student retrieval as well (Fig. 1).

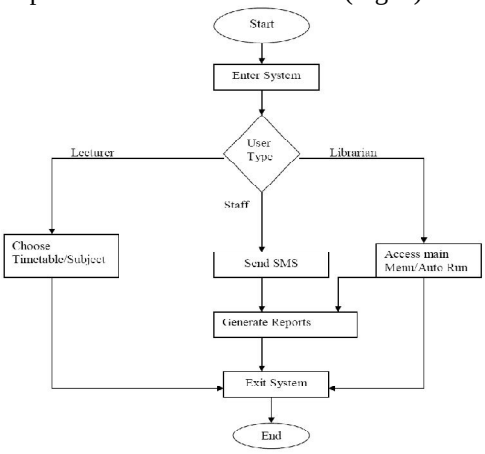


Figure 1: Proposed System Algorithm

There will be another sub-system in the library and in the records office. This system will have two options. The system in the records office will give the staff two options. One is to let the staff sort out which students have not paid their tuition fees yet and send reminders manually. The other one is the Autorun function. This function is about the system performing that task. The system will go through the student database and identify the students who have not settled their tuition fees yet. The system will then

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

prompt the user for approval before sending out pre-formatted notification emails or SMS to the sorted students (Fig. 2).

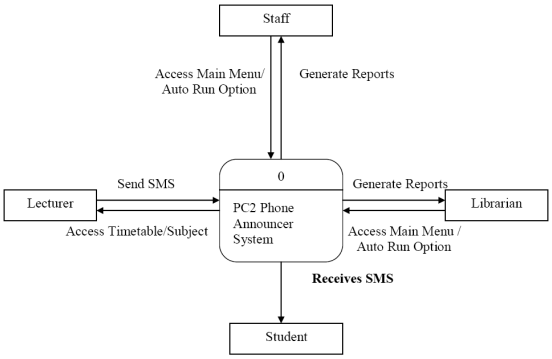


Figure 2: Proposed System Sub algorithm

A similar system will run in the library but the difference is that it will scan the database for overdue books and fines (Fig. 3).

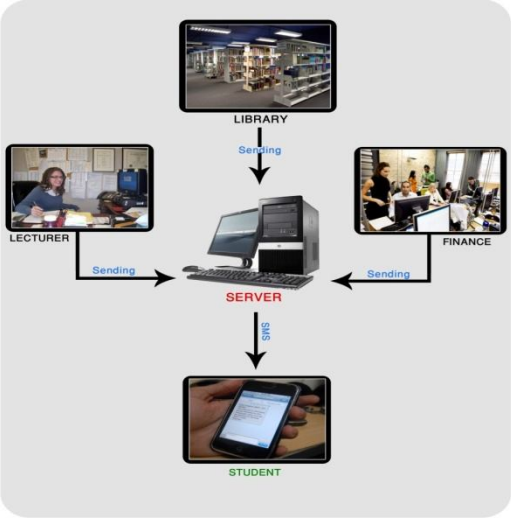


Figure 3: Proposed System Schema

CONCLUSION

Going through many articles on SMS technology as a means of communication to send bulk SMS, has helped in gaining knowledge about SMS technology usage in various organizations. We have gained knowledge about how SMS technology can be used as a means of communication between students, lecturers and management. Apart from SMS service being just as a means of communication how best the information available with individuals and organizations can be communicated to students in universities is conducted with a view to pass valuable information that benefits the students.

After analyzing the system and collecting the feedback from the lecturers, students on the existing system, we have come to this conclusion that PC 2 Phone Event Announcer is the best solution to overcome the existing communication problems in the Universities.

FUTURE WORK

Although the application of mobile phone technologies is not something new in the academic environments, but the provision of college (Lecturers, Records office and Library Staff) related services is still considered new.

More research should be done in this area where the concentration is put on the amount of information that needs to be filtered and delivered in this new environment. The challenge in research is the type of information services that the industries should deliver and find solutions to the versatile limitations existing in mobile phone technology in order to ensure satisfactory services to the users.

Firms need to work out the best ways to serve the users by the use this new mode of communication. They need to have constant collaboration with the researchers in the ICT fields, and also with the telecommunication providers to run a good study that looks into the appropriateness of industry functions that are to be delivered in the smaller interface environment.

The consequences of such study should be used in the actual execution of the services. In addition, the research conducted should also investigate the more advanced reference and information retrieval services and online public access with the help of the wireless technology.

ACKNOWLEDGMENT

We would like to express our sincere gratitude to Arash Habibi Lashkari, PHD candidate of UTM University for his supervision and guidance. Also, we would like to express our appreciation to our parents and all the teachers and lecturers who help us to understand the importance of knowledge and show us the best way to gain it.

REFERENCES

[1] Wonsang Song, Jong Yul Kim, Henning Schulzrinne,Dept. of Computer Science, Columbia University, and Piotr Boni and Michael Armstrong Verizon Network and Technology: Using IM and SMS for Emergency Text Communications.

[2] Malik Sikandar Hayat Khiyal, Aihab Khan, and Erum Shehzadi, Software Engineering Dept., Fatima Jinnah Women University, Rawalpindi, Pakistan(2009) SMS Based Wireless

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Home Appliance Control System (HACS) for Automating Appliances and Security.

[3] Fadhil T. Aula, University of Salahaddin, College of Engineering, Electrical Engineering Department Erbil, Iraq: Using SMS in Mobile Phone for Home Appliances Controlling

Through PC Parallel Port Interfacing.

[4] Daniel Mavrakis, MONACO TELEMATIQUE MC-TEL(2004). Integrate SMS in your enterprise

[5] V. Dinusha, S.M.K.D. Arunatileka,K.R.P. Chapman, G.P. Seneviratne, S. Saatviga, D. Wijethilake, S.Y.Y.D. Wickramasinghe," Enhancing e-Health using M-Communication in a Developing Country".

[6] Ashraf A. Tahat, School of Electrical Engineering, Princess Sumaya University for Technology, Amman 11941, Jordan (2009) MobileMessaging Services-Based Personal Electrocardiogram Monitoring System.

[7] Dr James Kadirire, : The short message service (SMS) for schools/conferences

http://www.formatex.org/micte2006/

[8] Tony Dwi Susanto, Informatics Engineering Department, Universitas Pembangunan Nasional "Veteran" Yogyakarta, Indonesia, &Robert Goodwin, School of Informatics and Engineering, The Flinders University of South Australia (2006): IMPLEMENTING SMS-BASED e-GOVERNMENT MODEL.

[9] Michal Yerushalmy & Oshrat Ben-Zaken, The Institute for Alternatives in Education, University of Haifa(2004), Mobile phones in Education: the case of mathematics

[10] Saranphong Pramsane, Ph.D. Graduate School of Information Technology, Assumption University, Bangkok, Thailand & Ridwan Sanjaya, Soegijapranata Catholic University, Semarang, Indonesia (2006) Mobile Education Services Based on SMS and Their Architecture Comparison.

[11] Awodele,Babcock University, Ilishan-Remo, Ogun State, Nigeria& E. R. Adagunodo Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria, A. T. ,Akinwale University of Agriculture, Abeokuta, Ogun State, Nigeria, & S. Idowu and M. Agbaje-Babcock University, Ilishan-Remo,Ogun State, Nigeria (2009). An Improved SMS User InterfaceResult Checking System

[12] Didier Chincholle, Michael Björn, Cristian Norlin and Morgan Lindqvist (2008) Chat on a phone, not a PC clone: IMS-based mobile community service.

[13] Rajiv Chakravorty, Student Member, IEEE, Sachin Katti, Student Member, IEEE, Ian Pratt, Member, IEEE, and Jon Crowcroft, Fellow, IEEE (2005): Using TCP Flow-Aggregation to Enhance Data Experience of Cellular Wireless Users.

[14] Laurie Butgereit, Meraka Institute, CSIR, Pretoria, South Africa, lbutgereit (at) meraka.org.za: Using Instant Messaging over GPRS to help with school work

[15] Neseem Al-Rawi, University of Portsmouth, Department of Information Science, Computer Event Communication via Mobile SMS (Short Message Service).

[16] André N. Klingsheim, Department of Informatics, University of Bergen(2004): Bluetooth and J2ME to Send and Receive SMS from Pc To Phone.

http://www.pdfone.com/keyword/controlling-pc-via-mobile-bluetooth-in-j2me.html

[17]KLINKMANN: GSM Control SMS Gateway, www.klinkmann.com

[18] Gogineni, M. and A. Ratan. "Evaluating the viability of a mobile phone-based, SMS/GPRS-enabled, client data collection channel for urban microfinance." Microsoft Research India Case Study, May 2007.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A Review on Ontology-Driven Query-Centric Approach for INDUS Framework

L. Senthilvadivu, Dept of Software Technology
SSM College of Engineering
Komarapalayam, Tamilnadu, India
lsvadivu.ssm@gmail.com

Dr. K. Duraiswamy, Dean(Academic)
K.S.R College of Technology
Tiruchengode, Tamilnadu, India
drkduraiswamy@yahoo.co.in

Abstract- This paper stimulates and describes the data integration component of INDUS that is, **In**telligent **D**ata **U**nderstanding **S**ystem, environment for data-driven information extraction and integration from heterogeneous, distributed, autonomous information sources. INDUS employs ontologies and inter-ontology mappings, to enable a user or an application to view a collection of physically distributed autonomous, semantically heterogeneous data sources regardless of location, internal structure and query interfaces as though they were a collection of tables structured according to an ontology supplied by the user. This allows INDUS to answer user queries against distributed, semantically heterogeneous data sources without the need for a centralized data warehouse or a common global ontology. The design of INDUS is motivated by the requirements of applications such as scientific discovery, in which it is desirable for users to be able to access, flexibly interpret, and analyze data from diverse sources from different perspectives in different contexts. INDUS implements a federated, query-centric approach to data integration using user-specified ontologies. More than 13 systems are studied and it is realized that INDUS is the most preferred system for Information Extraction, Integration, and Knowledge Acquisition from Heterogeneous, Distributed and Autonomous Information Sources. PROSITE, MEROPS, SWISSPROT, and MEME are examples of data sources used by Computational Biologists.

Keywords- INDUS (Intelligent Data Understanding System), Query-centric approach, PROSITE, MEROPS, SWISSPROT, MEME, MIPS2GO, EC2GO.

I. INTRODUCTION

INDUS is a modular, extensible, platform which does not dependent environment for information integration and data-driven knowledge acquisition from heterogeneous, distributed, autonomous information sources. INDUS when compared with machine learning algorithms for ontology-guided knowledge acquisition that can accelerate the pace of discovery in emerging data-rich domains such as biological sciences, atmospheric sciences, economics, defense, social sciences, by means of enabling scientists and decision makers rapidly and flexibly explore and analyze vast amounts of data from disparate sources. IBM provides a family of data management products that enable a systematic approach to solve the information integration challenges that businesses face today. Data Integration systems [2] attempt to provide users with seamless and flexible access to information from multiple autonomous, distributed and heterogeneous data sources through a unified query interface. Ideally, a data integration system should allow users to specify what information is needed without having to provide detailed instructions on how or from where to obtain the information. Data integration system must provide mechanisms for the following, such as communications and interaction with each data source as needed, specification of a query, expressed in terms of a user specified vocabulary, across multiple heterogeneous and autonomous data sources, specification of mappings between user ontology and the data-source specific ontologies, transformation of a query into a plan for extracting the needed information by interacting with the relevant data sources, and integration and presentation of the results in terms of a vocabulary known to the user. Basically there are two broad classes of approaches to data integration: Data Warehousing and Database Federation [4].

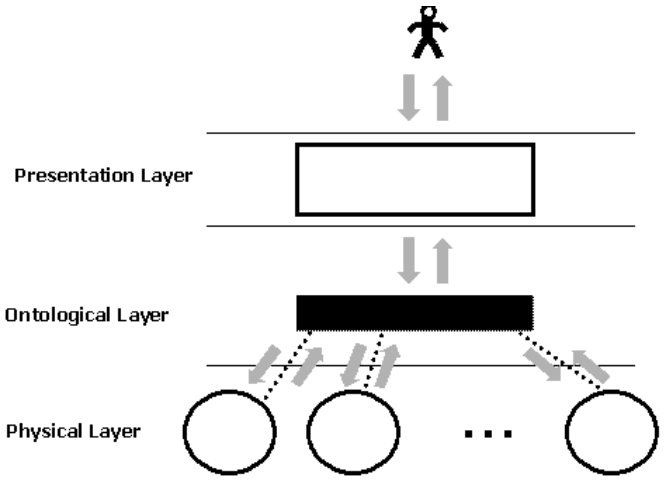


Figure1 Data Integration Layer

INDUS allows users to,

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

- View the set of data sources as if they were located locally and they were using a homogeneous interface.
- Interact with data sources (i.e., posting queries) through a provided interface that takes advantage of the functionality offered by each of data source using the query capabilities offered by the data sources to answer queries.
- Define their own language for defining queries and receiving answers.
- Define new concepts based on other concepts by applying a set of well-defined compositional operations.
- Use different definitions for the same concept, facilitating the exploration of new paradigms that explain the world.

For information integration and extraction from heterogeneous, distributed multi-relational information sources, this has implications in terms of how new basic concepts are incorporated into the system. Consider a system in which the query language is restricted to set union operations applied over EDI predicates without built-in predicates. Assuming a Query-centric case, Figure 2 shows a family of queries Q_i based on a set of basic concepts qij. Let I (Q) and I'(Q) be the set of instances satisfying Q respectively before and after adding c to the system. Assume that $\forall_i$ I(c) ≠ I (Q_i). Then, $\forall_i$ such that c is added to G (Q_i), I (Q_i) ≠ I'(Q_i). In other words, only those queries where c is explicitly added may return a different answer.

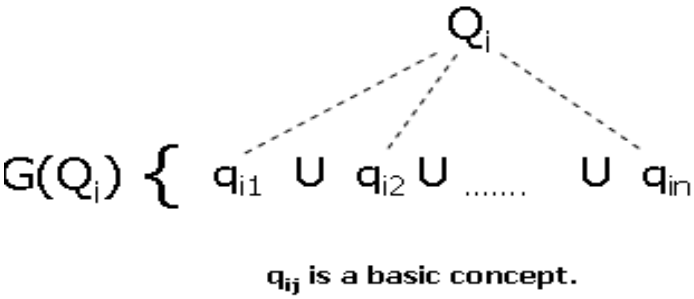


Figure 2 Query-Centric Approach Examples.

Data sources are autonomous, distributed, and heterogeneous in structure and content; the complexity associated with accessing the data answering queries must be hidden from the users; the users need to be able to view disparate data sources from their own point of view. INDUS consists of three principal layers. In the lower part, the set of data sources accessible by INDUS are shown. In the physical layer, a set of instantiators enable INDUS to communicate with the data sources. The ontological layer offers a repository where ontologies are stored. Using these repository syntactical and semantic heterogeneities may be solved. Also, another relational database system is used to implement the user workspace private area where users materialize their queries. The user interface layer enables users to interact with the system.

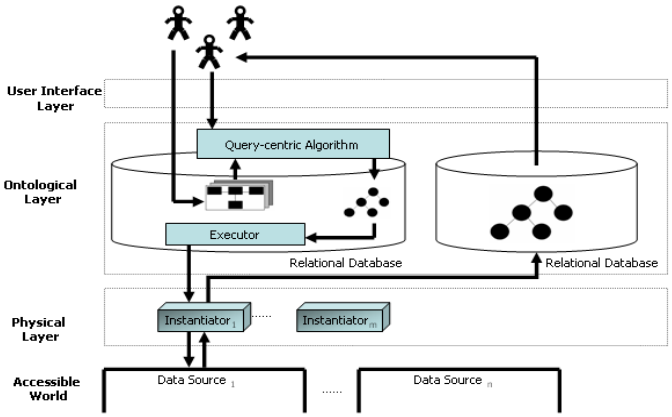


Figure3 INDUS Schematic Diagram

INDUS is based on five modules. The graphical user interface enables users to interact with INDUS. This module is developed under Oracle Developer 6i. The common global ontology area, implemented through a relational database system, stores all information about ontologies, concepts and queries. Any information stored in this repository is shared for all users. The private workspace user area is also implemented through a relational database system. Each INDUS user has a private area where queries are materialized.

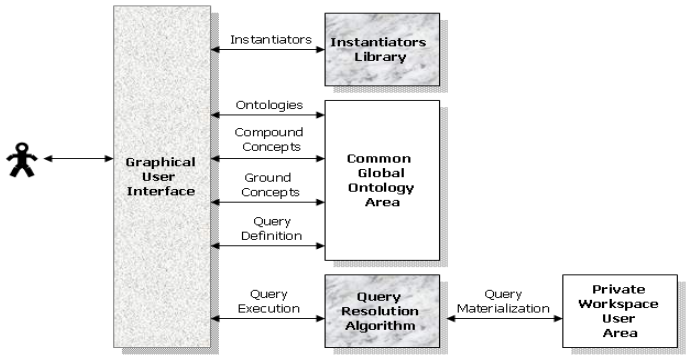


Figure 4 INDUS Module Diagram

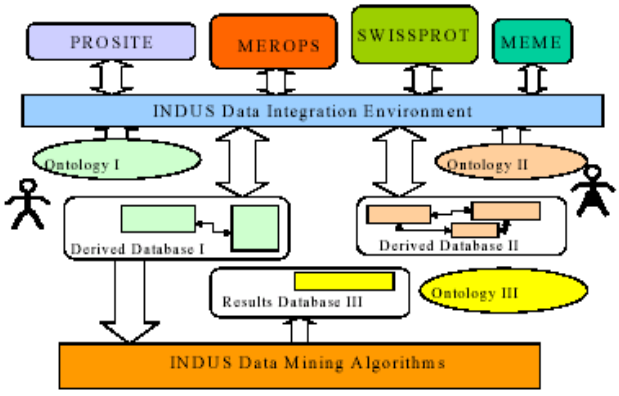


Figure5 INDUS

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The rest of the paper is organized as follows: Section II briefly introduces the related work done by various authors and section III conclude and enhance the future work of INDUS.

II. RELATED WORK

Sudarshan Chawathe et al., stated the main motive of the Tsimmis Project, this is mainly to develop tools that facilitate the rapid integration of heterogeneous information sources that may include both structured and unstructured data. This paper gives an overview of the project, describing components that extract properties from unstructured objects, which translate information into a common object model, that combine information from several sources, that allow browsing of information, and that manage constraints across heterogeneous sites. Tsimmis is a joint project between Stanford and the IBM Almaden Research Center. In summary, the Tsimmis project is exploring technology for integrating heterogeneous information sources. Current efforts are focusing on translator and mediator generators, which should significantly reduce the effort required to access new sources and integrate information in different ways. TSIMMIS architecture is based on the concept of wrappers and mediators. Each wrapper knows how to deal with a particular data source and it is able to receive a query in a common language - Object Exchange Model (OEM) and to transform it into a particular language understood by the data sources. Both INDUS and TSIMMIS use query-centric approach to data integration. However, unlike TSIMMIS, INDUS maintains a clear separation between ontologies used for data integration (which are supplied by users) and the procedures that use ontologies to perform data integration. This allows INDUS users to replace ontologies used for data integration 'on the fly'. This makes INDUS attractive for data integration tasks that arise in exploratory data analysis wherein scientists might want to experiment with alternative ontologies.

Pegasus [17], a heterogeneous multi-database management system that responds to the need for effective access and management of shared data across in a wide range of applications. Pegasus provides facilities for multi-database applications to access and manipulate multipole autonomous heterogeneous distributed object-oriented relational and other information systems through the uniform interface. It is a complete data management system that integrates various native and local databases. Pegasus takes advantage of object-oriented data modeling and programming capabilities. It uses both type and function abstractions to deal with mapping and integration problems. Function implementation can be defined in an underlying database language or a programming language. Data abstraction and encapsulation facilities in the Pegasus object model provide an extensible framework for dealing with various kinds of heterogeneities in the traditional database systems and nontraditional data sources. UniSQL/M [18], [19], SIMS [20], IRO-DB [3], and other projects, support mediator capabilities through a unified global schema [21], which integrates each remote database and resolves conflicts among these remote databases. Although these projects made substantial contributions in resolving conflicts among different schemas and data models, the global schema approach suffers from the fragile mediator problem; the unified global schema must be substantially modified as new sources are integrated. For example, UniSQL/M [18], [19] is a commercial multi-database product; virtual classes are created in the unified schema to resolve and "homogenize" heterogeneous entities from relational and object-oriented schema. Instances of the local schema are imported to populate the virtual classes of the integrated schema, and this involves creating new instances. The first step in integration is defining the attributes of a virtual class, and the second step is a set of queries to populate this class. They provide a vertical join operator, similar to a tuple constructor, and a horizontal join, which is equivalent to performing a union of tuples. The major focus of their research conflicts due to generalization, for e.g., an entity in one schema can be included, i.e., become a subclass of an entity in the global schema, or a class and its subclasses may be included by an entity in the global schema. Attribute inclusion conflicts between two entities can be solved by creating a subclass relationship among the entities. Other problems that are studied are aggregation and composition conflicts. Alternately, the capability of a mediator to resolve conflicts is supported by the use of higher-order query languages or meta-models [22], [23], [24]. Mediators are also implemented through the use of mapping knowledge bases that capture the knowledge required to resolve conflicts among the local schema, and mapping or transformation algorithms that support query mediation and interoperation among relational and object databases.

Jaime A Reinoso Castillo motivates and describes the data integration component of INDUS (Intelligent Data Understanding System) environment for data-driven information extraction and integration from heterogeneous, distributed, autonomous information sources. The design of INDUS is motivated by the requirements of applications such as scientific discovery, in which it is desirable for users to be able to access, flexibly interpret, and analyze data from diverse sources from different perspectives in different contexts. INDUS implements a federated, query-centric approach to data integration using user-specified ontologies. Development of high throughput data acquisition in a number of domains (e.g. biological sciences, space sciences, commerce) along with advances in digital storage, computing, and communication technologies have resulted in unprecedented opportunities in data-driven knowledge acquisition and decision making. The effective use of increasing amounts of data from disparate information sources presents several challenges in practice. This paper describes the data integration component of INDUS (Intelligent Data Understanding System) – a modular, extensible, platform independent environment for information integration and data-driven knowledge acquisition from heterogeneous, distributed, autonomous information sources. INDUS when equipped with

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

machine learning algorithms for ontology-guided knowledge acquisition can accelerate the pace of discovery in emerging data-rich domains (e.g., biological sciences, atmospheric sciences, economics, defense, social sciences) by enabling scientists and decision makers to rapidly and flexibly explore and analyze vast amounts of data from disparate sources.

Paton N.W, et al., described about the Transparent Access to Multiple Bioinformatics Information System (TAMBIS) is an ontology centered system for evaluating queries that offers access to multiple heterogeneous bioinformatics data sources. TAMBIS is based on three-layer wrapper/mediator architecture. Like INDUS, it uses a query-centric approach to data integration. It includes an ontological layer and a graphical user interface for querying. The ontology allows the creation of new concepts based on compositional operations of previously defined concepts using a restricted grammar based on the description logic language GRAIL. TAMBIS returns the answer for a query as an HTML file. Thus, the size of the main memory may limit the amount of data that may be returned in response to a query. In contrast, INDUS stores the answer for a query in a user private area implemented by a relational database system. Thus, queries that return large amounts of data are manipulated more efficiently in terms of hardware and software resources. INDUS also provides better support for defining multiple ontologies for use in different contexts by different users.

Arnaud Sahuguet et al., put forth an introduction to WysiWyg Web Wrapper Factory (W4F) a toolkit for generating Web wrappers. It contains a language for identifying and navigating Web sites with retrieval rules and a declarative language for extracting data from Web pages with extraction rules. It also provides a mechanism for mapping extracted data to a target structure. As its name suggests, W4F provides a graphical user interface for generating retrieval, extraction, integration and mapping rules. While W4F and ANDES are similar in many respects, their main difference is that whereas W4F uses a proprietary language for data extraction and mapping rules, ANDES is based on XHTML and XSLT and can exploit templates, recursive path expressions, and regular expressions for more effective data extraction, mapping, integration and aggregation. Hyperlink synthesis, which allows data to be extracted from the "deep Web," is also unique to ANDES. The goal of WIDL is to define a programmatic interface to Web sites. As such, it focuses more on the mechanics of how to issue a request to a Web site, retrieve the result, and bind the input and output variables to a host programming language, than the process of extracting data from the retrieved result page. WIDL allows data to be extracted using absolute path expressions, this falls short of building robust data extractors. Feature extraction and structure synthesis would be difficult to implement in WIDL and would be relegated to some higher-level program. The Web Language (formerly WebL) from Compaq is a procedural language for writing Web wrappers and it provides a powerful data extraction language, which is similar to recursive path expressions combined with regular expressions and the language is not tuned to XML inputs and outputs and lacks the power of XSLT templates and XPath axes and operators.

Naveen Ashish et al., had a look on three systems based on information extraction and integration, they are Ariadne, Garlic, and TSIMM systems, which are seems to be mediators that facilitate querying multiple heterogeneous sources. While Garlic and TSIMMIS support a wide range of sources, including Web sources, database systems, and file systems, Ariadne focuses on Web sources exclusively. In each system, a modeling process produces an integrated view of the data contained in the sources and a query planning process decomposes queries on the integrated view into a set of subqueries on the sources. In Garlic and TSIMMIS, wrappers are written in a procedural programming language and are compiled into executable code, whereas in Ariadne, an induction-based wrapper generation mechanism is used. It uses regular expressions and includes mapping tables to resolve vocabulary differences between Web sources, but lacks path expressions. The path expressions are important in extracting data from a HTML tree because hierarchical navigation between nested HTML elements is frequently needed. In ANDES, a combination of XPath axes and operators with regular expressions provides for more robust data extraction rules than what is possible with regular expressions alone.

An early study on biological data integration was done by Marcotte et al., gave a combined algorithm for protein function prediction based on microarray and phylogeny data, by classifying the genes of the two different datasets separately, and then combining the genes' pair-wise information into a single data set. The approach does not scale immediately. The method extends to a general combinatorial data integration framework based on pair-wise relationships between elements and any number of experiments. In machine learning, Pavlidis et al. use a Support Vector Machine algorithm to integrate similar data sets in order to predict gene functional classification. The methods use a lot of hand tuning with any particular type of data both prior and during the integration for best results. Troyanskaya et al. use a Bayesian framework to integrate different types of genomic data in yeast. The author's probabilistic approach is parallel to the combinatorial approaches. A lot of work has been done on specific versions of the consensus clustering problem, based on the choice of a distance measure between the clustering's and the optimization criterion. Strehl et al., use a clustering distance function derived from information theoretic concepts of shared information. Recently, Monti et al. used consensus clustering as a method for better clustering and class discovery. Other authors have also used the quota rule in the past. Finally, Cristofo and Simovici have used Genetic Algorithms as a heuristic to find median partitions. The author shows that the approach does better than several others among

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

which a simple element move algorithm, which coincidently the algorithm has also been shown to be better than recently. It would be interesting in the near future to compare the machine learning methods with the combinatorial approach.

Witold Abramowicz et al., gave the importance of Deep Web (DW) has grown substantially in recent years not only because its size, but also because Deep Web sources arguably contain the most valuable data, as compared to the so-called Surface Web. An overlap analysis between pairs of search engines conducted in 2001 has estimated that there exist ca. 200,000 Deep Web sites, providing access to 7,500TB of data. Further studies (2004) have estimated the number of Deep Web sites to reach slightly more than 300,000, providing access to ca. 450,000 databases through 1,260,000 query interfaces. A lot of research in previous years has been devoted to information extraction from the Web (IEW). Data integration (DI) problems including schema mapping, query capabilities description and data translation and cleaning were studied in extend in previous years. Today, these research areas converge, leading to development of systems for Deep Web data extraction and integration (DWI). Deep Web poses new challenges to data extraction as compared with Surface Web. New problems arise also for data integration unknown in traditional databases. This paper identifies 13 systems most prominently referred in the subject literature, they are AURORA, DIASPORA, Protoplasm, MIKS, TSIMMIS, MOMIS, GARLIC, SIMS, Information Manifold, Infomaster, DWDI, PICSEL, and Denodo to base the architecture on the approaches reported to date.

Chantal Reynaud et al., proposed with integration of XML heterogeneous information sources into a data warehouse with data defined in terms of a global abstract schema or ontology. The authors present an approach supporting the acquisition of data from a set of external sources available for an application of interest including data extraction, data transformation and data integration or reconciliation. The integration middleware that the authors propose extracts data from external XML sources which are relevant according to RDFS+ ontology, transforms returned XML data into RDF facts conformed to the ontology and reconciles RDF data in order to resolve possible redundancies. RDFS+ can be viewed as a fragment of the relational model which is restricted to unary and binary relations enriched with typing constraints, inclusion and exclusion between relations and functional dependencies. Data extraction and transformation are completely automatic tasks usually performed by wrappers. It is a two-step process. First, an abstract description of the content of the external source is built. Second, data is extracted and presented in the form of the data warehouse. This paper has presented an information integration approach able to extract, transform and integrate data in a data warehouse guided by ontology. This approach can be applied to XML sources that are valid documents and that have to be integrated in a RDF data warehouse with data described in terms of RDFS ontology. Mappings between the external sources and the ontology are represented in a declarative way. Extraction operates on any XML document given mappings represented in XPath in terms of the ontology. Data transformation consists in converting data in terms of the ontology and in the same format.17 Data Extraction, Transformation and Integration Guided by Ontology. Both tasks are performed through XML queries associated to views of the sources automatically built beforehand. Through data integration, the author addressed the reference reconciliation problem and presented a combination of a logical and numerical approach. Both approaches exploit schema and data knowledge given in a declarative way by a set of constraints and are then generic. The relations between references are exploited either by L2R for propagating non-reconciliation decisions through logical rules or by N2R for propagating similarity scores thanks to the resolution of the equation system. The two methods are unsupervised because no labeled data set is used. Furthermore, the combined approach is able to capitalize its experience by saving inferred synonymies. The results that are obtained by the logical method are sure. This distinguishes L2R from other existing works. The numerical method complements the results of the logical one. It exploits the schema and data knowledge and expresses the similarity computation in a non linear equation system. The experiments show promising results for recall, and most importantly its significant increasing when constraints are added.

Cui Tao had a new thought of information extraction and integration from heterogeneous biological data sources. This paper deals about the huge and growing amounts of biological data that reside in different online repositories. Most of these Web-based sources only focus on some specific areas or only allow limited types of user queries. To obtain needed information, biologists usually have to traverse different Web sources and combine their data manually. In this research, author proposes a system that can help users to overcome these difficulties. Given a user's query within the area of molecular biology, the system can automatically discover appropriate repositories, retrieve useful information from these repositories and integrate the retrieved information together.

Aditya Telang et al., projected about information integration across heterogeneous sources. In this survey paper, it identifies the set of challenges that need to be addressed for this form of heterogeneous information integration, and compare the current state-of-the-art as to how they fare. This paper proposes a framework with functional components - termed InfoMosaic, which aims to address some of these important challenges, and briefly elaborate on the data and control low involved in answering a complex query/search.
As more and more data becomes available on the web, it is even more important to be able to search for complex queries instead of humans performing the task of information integration using basic search capabilities. Indeed, the use of Web needs to move towards more specialized content-based

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

retrieval mechanisms (such as information integration) that do more than simply return documents. Extensive work is needed on the higher-levels of the system, including managing semantic heterogeneity in a more scalable fashion, the use of domain knowledge in various parts of the system, transforming these systems from query-only tools to more active data sharing scenarios, and easy management of data integration systems. Extensibility of the system and the framework is extremely important as the coverage of the system should increase as we add more domain knowledge and source semantics. The objective of InfoMosaic is to allow users to specify what information is to be retrieved without having to provide detailed instructions on how or from where to obtain this information. This approach draws upon techniques from database systems, artificial intelligence, information retrieval, and the use of extended ontologies.

C.A. Knoblock et al., revised about, wrappers are typically employed by most frameworks for the extraction of heterogeneous data. However, as the number of data sources on the web and the diversity in their representation format continues to grow at a rapid rate, manual construction of wrappers proves to be an expensive task. There is a rapid need for developing automation tools that can design, develop and maintain wrappers effectively. Even though a number of integration systems have focused on automated wrapper generation such as Ariadne's Stalker, MetaQuerier, TSIMMIS, InfoMaster, and Tukwila, since the domains embedded within these systems are known and predefined; the task of generating automated wrappers using mining and learning techniques is simplified by a large extent. There also exist several independent tools based on solid formal foundations that focus on low-level data extraction from autonomous sources such as Lixto, Stalker, etc… In the case of spatial data integration, ontologies and semantic web-services are defined for integrating spatial objects, in addition to wrappers and mediators. Heracles combines online and geo-spatial data in a single integrated framework for assisting travel arrangement and integrating world events in a common interface. A Storage Resource Broker was proposed in the LTER spatial data workbench to organize data and services for handling distributed datasets. Information Manifold claimed that the problem of wrapping semi-structured sources would be irrelevant as XML will eliminate the need for wrapper construction tools. This is an optimistic yet unrealistic assumption since there are some problems in querying semi-structured data that will not disappear, for several reasons: 1) some data applications may not want to actively share their data with anyone who can access their web-page, 2) legacy web applications will continue to exist for many years to come, and 3) within individual domains, XML will greatly simplify the access to sources; however, across diverse domains, it is highly unlikely that an agreement on the granularity for modeling the information will be established.

Paul Buitelaar et al., deliberately shows the idea of ontology-based information extraction with SOBA. This paper describes SOBA, a sub-component of the SmartWeb multi-modal dialog system. SOBA is a component for ontology based information extraction from soccer web pages for automatic population of a knowledge base that can be used for domain specific question answering. SOBA realizes a tight connection between the ontology, knowledge base and the information extraction component. The originality of SOBA is in the fact that it extracts information from heterogeneous sources such as tabular structures, text and image captions in a semantically integrated way. In particular, it stores extracted information in a knowledge base, and in turn uses the knowledge base to interpret and link newly extracted information with respect to already existing entities. SmartWeb is a multi-modal dialog system that derives answers from unstructured resources such as the Web, from automatically acquired knowledge bases and from semantic web services. The extracted information is defined with respect to an underlying ontology (SWIntO): SmartWeb Integrated Ontology to enable a smooth integration of derived facts into the general Smart-Web system. Ontologically described information is a basic requirement for more complex processing tasks such as reasoning and discourse analysis. More in particular, there are three main reasons for formalizing extracted information with respect to ontology. This paper described SOBA, an information extraction system which relies on ontology to formalize and semantically integrate extracted information from heterogeneous resources in a knowledge base.

Hicham snoussi et al, established the heterogeneous web data extraction using ontology. Multi-agent systems can be fully developed only when they have access to a large number of information sources. These latter are becoming more and more available on the Internet in form of web pages. This paper does not deal with the problem of information retrieval, but rather the extraction of data from HTML web pages in order to make them usable by autonomous agents. This problem is not trivial because of the heterogeneity of web pages. Users and agents can query the extracted data using a standard querying interface. The ultimate goal of this tool is to provide useful information to autonomous agents. This approach does not rely on identification of boundaries of character strings within HTML documents, as is the case in TSIMMIS. In this paper, the author dealt with the problem of data extraction from web pages and their integration in applications. In particularly the goal of this paper is to find a way to extract reliable data, and to convert them in a standard form. The extraction of data consisted in two steps: converting an HTML page into XML and using XQL to query XML documents to extract the desired data. The extraction process is controlled by a specification file, which describes what elements of a web page to extract, and how they have to be extracted. As the user has a tight control on the extraction process, the extracted data are of high quality, thus can be exploited by other programs or software agents.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The integration of data is based on the use of an ontology, which provides a common model for information sources. All the data extracted fit the conditions of the ontology, which makes the data integration easier. The use of ontology has greatly simplified the task of extraction and integration. The most critical point remained is the definition of an ontology. However, we cannot imagine an open system that exchanges data without using a norm of the domain. Even if one cannot construct a complete ontology, a standard will always be necessary to play a similar role. The advantage of such a specification is that, once constructed, it can be reused for similar applications. Moreover, the same specification can be exploited by software agents to get data. The most useful case of such an extraction process is on web pages that present dynamic contents, but with fixed structures. Examples are web pages that provide stock market exchange prices, money exchange rates, and so on. If an information site is restructured, the extraction process is no longer valid.

III. Conclusion And Future Enhancement

In this paper, we have described and surveyed the design and implementation of the systems which integrate the heterogeneous information sources. The data integration component of INDUS (Intelligent Data Understanding System) environment for flexible information extraction and information integration and knowledge acquisition from heterogeneous, distributed, autonomous information sources is more suitable system. INDUS implements a federated, query-centric approach to data integration. Hence, the information extraction operations to be executed are dynamically determined on the basis of the user-supplied ontology and the query supplied by the user or an application program. The approach has been applied successfully to scenarios where the ontologies associated with some attributes are given by tree structured hierarchies. It is desirable to extend the work to the more general case where the hierarchies are directed acyclic graphs, as this case is more often encountered in practice. As Protege is the most popular tool for creating knowledge bases, in the future INDUS will allow users to import ontologies that are edited using Protege. It is of interest to extend INDUS to scenarios where each data sources can be conceptually viewed as a set of inter-related possibly hierarchical tables. This requires a framework for asserting semantic correspondences between tables and relations across multiple ontologies. In this context, recent work on description logics for representing and reasoning with ontologies, distributed description logics as well as ontology languages, e.g., web ontology language (OWL) are of interest. These developments, together with the work on INDUS, set the stage for making progress on the problem of integration of a collection of semantically heterogeneous data sources where each data source can be conceptually viewed as a set of inter-related tables in its full generality.

References

[1] Honavar, V., Millar, L., and Wong, J. 1998. Distributed Knowledge Networks. Design, Implementation, and Applications In: Proceedings of the IEEE Information Technology Conference. pp. 87-90. IEEE Press.

[2] Calvanese, D., Giacomo, G., Lenzerini, M., et al., 1998. Information integration: Conceptual modeling and reasoning support. In: CoopIS'98 [3] Levy, A., 1998. The Information Manifold approach to data integration. In: IEEE Intelligent Systems, 13 12-16, August 19 1998.

[3] Haas L.M., Schwarz, P.M., Kodali, P., Kotlar, E., Rice, J.E., Swope, W.P., 2001. DiscoveryLink: A system for integrated access to life sciences data sources. In: IBM System JThenal . Vol 40. No 2. 2001.

[4] Levy, A., 2000. Logic-based techniques in data integration. Logic Based Artificial Intelligence, Edited by Jack Minker. Kluwer Publishers.

[5] Ullman, J., 1997. Information integration using logical views. In: 6th ICDT.Pages 19-40, Delphi, Greece.

[6] Garcia-Molina , Y. Papakonstantinou , D. Quass , A. Rajaraman , Y. Sagiv , J. Ullman , V. Vassalos , J. Widom (1996). The TSIMMIS approach to mediation: Data models and Languages. *JThenal of Intelligent Information Systems*

[7] Paton, N.W., Stevens, R., Baker, P.G., Goble, C.A., Bechhofer, S., 1999. Query processing in the TAMBIS bioinformatics source integration system. In: Proc. 11th Int. Conf. on Scientific and Statistical Databases (SSDBM), IEEE Press, 138-147, 1999.

[8] Stevens, R., Baker, P., Bechhofer, S., Ng, G., Jacoby, A., Paton, N., Goble, C., Brass, A, 2000. TAMBIS: Transparent access to multiple bioinformatics information sources. Bioinformatics, 16:2 PP.184-186.

[9] Reinoso Castillo, J 2002. Ontology-Driven Information Extraction and Integration from Autonomous, Heterogeneous, Distributed data sources – A Federated Query-Centric approach. Masters Thesis. Artificial Intelligence Research Laboratory, Department of Computer Science, Iowa State University.

[10] Wang, X., Schroeder, D., Dobbs, D., and Honavar, V. (2003). Data-Driven Discovery of Rules for Protein Function Classification Based on Sequence Motifs. Information Sciences. In press.

[11] Andorf, C., Dobbs, D., and Honavar, V. (2003) Reduced Alphabet Representations of Amino Acid Sequences for Protein Function Classification. Information Sciences. In press.

[12] Wong, J., Helmer, G., Naganathan, V. Polavarapu, S., Honavar, V., and Miller, L. (2001) SMART Mobile Agent Facility. JThenal of Systems and Software. Vol. 56. pp. 9-22.

[13] Caragea, D., Silvescu, A., and Honavar, V. (ISDA 2003) Decision Tree Induction from Distributed, Heterogeneous, Autonomous Data Sources. In: Proceedings of the Conference on Intelligent Systems Design and Applications.

[14] Caragea, D., Silvescu, A., and Honavar, V. (2001a). Analysis and Synthesis of Agents that Learn from Distributed Dynamic Data Sources. Invited chapter. In*:* Wermter, S., Willshaw, D., and Austin, J. (Ed.). *Emerging Neural Architectures Based on Neuroscience.* Springer-Verlag.

[15] Atramentov, A., Leiva, H., and Honavar, V. (ILP 2003). Learning Decision Trees from Multi-Relational Data. In: Proceedings of the Conference on Inductive Logic Programming To appear.

[16] Arens,Y., Chee, C., Hsu, C., and Knoblock, C. (1993) Retrieving and Integrating Data from Multiple Information Sources. *International JThenal of Intelligent and Cooperative Information Systems*. Vol. 2, No. 2. Pp. 127-158

[17] Zhang, J. and Honavar, V. (2003). Learning Decision Tree Classifiers from Attribute-Value Taxonomies and Partially Specified Data. In: Proceedings of the International Conference on Machine Learning. Washington, DC. In press.

[18] Madhavan, J., Bernstein, P., Halevy, A., and Domingos, P. 2002. Representing and Reasoning about Mappings between Domain Models. Proceedings of the Eighteenth National Conference on Artificial Intelligence(pp.80-86). Edmonton, Canada: AAAI Press.

[19] Draper, D., Halevy, A., and Weld, D. (2001). The NIMBLE XML Data integration System. In: Proceedings of the International Conference on Data Engineering (ICDE 01).

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[20] Sheth, A.P. and J. A. Larson (1990) Federated database systems for managing distributed, heterogeneous and autonomous databases, ACM Computing Surveys 22 pp. 183--236.
[21] Sowa, J. (1999) Knowledge Representation: Logical, Philosophical, and Computational Foundations. New York: PWS Publishing Co.
[22] Subrahmanian, V.S., Sibel Adali, Anne Brink, James J. Lu, Adil Rajput, Timothy J. Rogers, Robert Ross, Charles Ward (2000). HERMES A Heterogeneous Reasoning and Mediator System.
[23] Wiederhold, G. and M. Genesereth (1997) The Conceptual Basis for Mediation Services, IEEE Expert, Vol.12 No.5 pp. 38-47
[24] Yan, C., Dobbs, D., Honavar, V.,(2003) Identification of Residues Involved in Protein-Protein Interaction from amino acid sequence – A Support Vector Machine approach. In: Proceedings of Intelligent Systems Design and Application
[25] Knoblock, C.A., Minton, S. Ambite J.L., Ashish, N. Muslea, I. Philpot, A.G. and Tejada, S. The Ariadne Approach to Web-Based Information Integration. International the JThenal on Cooperative Information Systems 10(1/2), pp145-169, 2001.

Prof L.SenthilVadivu received M.Sc., M.Phil in Physics, PGDIT., MCA., M.Phil in Computer Science. She is currently working as Assistant Professor, in the Department of Software Technology in SSM College of Engineering, Komarapalayam. She has 16 years of experience in the field of Physics and Computer Science and shown her excellence by attending short term course in the field of Mobile Computing organized by K.S.R College of Engineering and a workshop in the area of Data Mining using Memetic Algorithm organized by Periyar University, Salem. She has published number of papers in national and international conferences and journals. She is a member of ISTE and IEEE. Her email is lsvadivu.ssm@gmail.com.

Dr.K. Duraiswamy B.E., M.Sc. (Engg.), Ph.D, MISTE, SMIEEE is currently working as Dean (Academic) in K.S. Rangasamy College of Technology, Tiruchengode – 637 215, Tamilnadu, India. He has 42 years of teaching and research experience. He has guided 16 Ph.Ds in the area of Computer Science and Engineering in addition to 14 M.Phil students in Computer Science. He is currently guiding more than 12 students for Ph.D. He has also guided more than 100 M.E. students in the area of Computer Science and Engineering. He has published 51 papers in International Journals and 12 papers in National Journals in addition to participating more than 72 National and 43 International Conferences. His area of interest are : Image Processing, Network Security, Data Communication, Soft Computing, State estimation, Power System load forecasting and scheduling, Computer Architecture character recognition Data mining etc. His email is drkduraiswamy@yahoo.co.in

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Fingerprint Image Enhancement using Successive Mean Quantization Transform

Dr.R.Seshadri ,B.Tech,M.E,Ph.D
Director, S.V.U.Computer Center
S.V.University, Tirupati
E-mail: ravalaseshadri@gmail.com

Yaswanth Kumar.Avulapati,M.C.A,M.Tech,(Ph.D)
Research Scholar, Dept of Computer Science
S.V.University, Tirupati
E-mail:yaswanthkumar_1817@yahoo.co.in

Abstract

Fingerprint based identification is one of the core methodologies for person identification. A finger print is the representation of the epidermis of a finger. In this paper we presents a novel approach to enhancing fingerprint images using a SMQT(Successive Mean Quantization Transform) for automatic enhancement of fingerprint gray scale images. There are so many approaches are there for image enhancement. Here we introduced the Fingerprint enhancement capabilities and properties of the transform are analyzed. The transform is capable to perform both a nonlinear and a shape preserving stretch of the fingerprint histogram.

Introduction

Fingerprint image quality is of much importance to achieve high performance in Automatic Fingerprint Identification System (AFIS). Several researches have proposed some enhancement techniques to this end.

Noise is an important factor that influences fingerprint quality which is mainly produced in the processes of fingerprint acquirement and transmission.

There is necessity for us to reduce noises in the fingerprint images to acquire useful data. The Successive Mean Quantization transform (SMQT) are satisfactory in stripped noise reduction. Noise reduction is necessary to do finger print image processing and fingerprint image interpretation so as to acquire useful information that we want. SMQT can be interpreted as a progressive focus on details in an image. These characteristics make the transform interesting for automatic enhancement of fingerprint image.

In this paper the SMQT (Successive Mean Quantization Transform) is applied for automatic fingerprint enhancement. The transform is presented using set theory. An adjustment parameter is introduced to further control the enhancement. The nonlinear properties of the transform are investigated by means of the histogram change. The image enhancement results are compared to histogram equalization.

Fingerprint Identification:

A finger print system works in two modes they are Enrollment mode and

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Authentication mode as shown in the fig.1. In the Enrollment mode fingerprint system is used to identify and collect the related information about the person and his/her fingerprint image. Authentication mode in which the fingerprint system is used to identify the person who is declared to be him/her as show

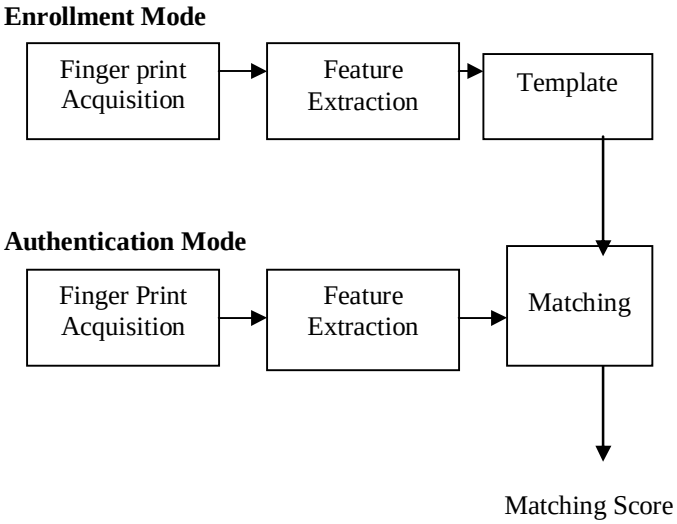


Fig 1.Enrollment and Authentication of a Fingerprint system

At the time of capturing the fingerprint images from the scanner some times it acquire noisy fingerprints .The performance of fingerprint identification system is depending on the quality of the fingerprint images due to the skin conditions like dry or wet ,burns, cuts and capturing sensor noises like manual labors a significant percentage of fingerprint images is of poor quality.

The goal of fingerprint enhancement algorithms is to improve the clarity of the fingerprint image (ridge structures) in the recoverable regions and unrecoverable regions as too noisy for further processing as shown in the fig 3(A). We proposed fingerprint image enhancement using Successive Mean Quantization Transform (SMQT).

FINGER PRINT IMAGE ENHANCEMENT

SMQT (Successive Mean Quantization Transform) is a straight forward way to enhance fingerprint image that can use directly. The only parameter to adjust is the level

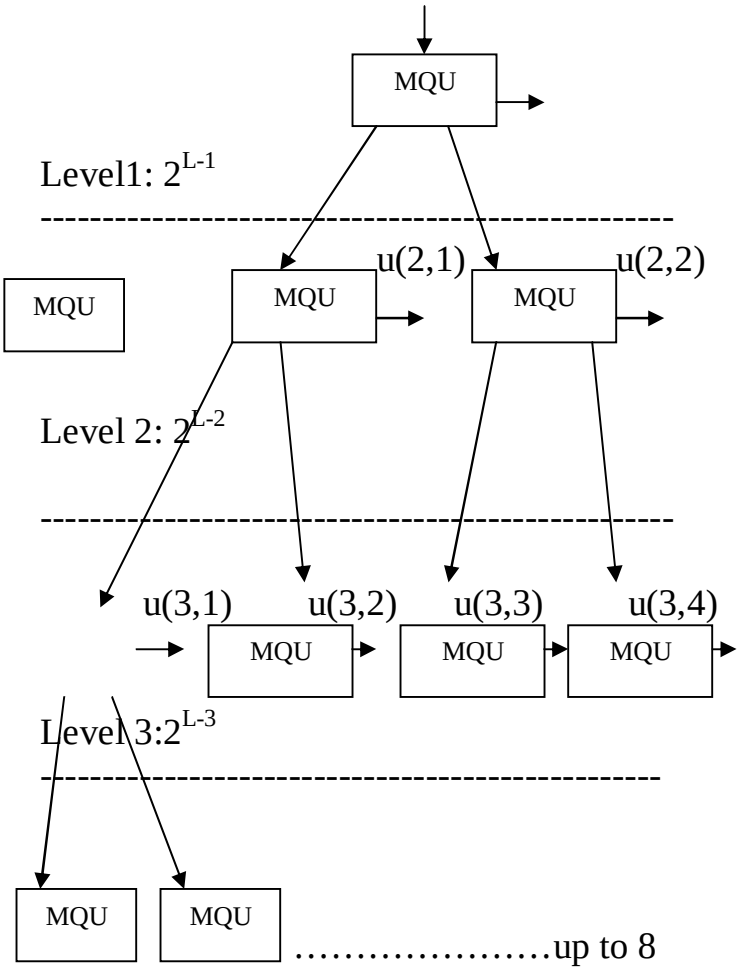


Fig 2.The SMQT as a binary tree of MQUs

The Successive Mean Quantization Transform (SMQT) as a binary tree of Mean Quantization Units (MQUs) L as shown in the fig.2. Today digital imaging devices typically use the range 0 . . . 255, that is 8 bits is used. For automatic Fingerprint image enhancement of 8 bits images L is chosen to 8. Nevertheless, it could be convenient to control the amount of enhancement applied. Given the original pixel set $D(x)$ and the SMQT8 enhanced pixel set$M(x)$, a simple mix of the two sets can be found as

$$T(z) = \{z \mid \mathbf{V}(z) = (1-\alpha)\mathbf{V}(x) + \alpha\mathbf{V}(y), \; x \; D, y \; M\}$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Where α is the blending factor. For default enhancement is $\alpha = 1$ chosen, for no enhancement is $\alpha = 0$ chosen and values in between controls the amount of blending. This type of control could for example be convenient for a human interpreting digital X-rays. However, $\alpha = 1$ is the default setting. Consider the enhancement of fingerprint found in Fig. 3(B). Note the contrast enhancement achieved by the SMQT. Investigations of the properties for the enhancement can be performed by analyzing the histogram

Fig. 3. (A) Original Fingerprint.

Fig.3 (B) SMQT8 Enhanced Fingerprint.

In the original image histogram it is possible to see that this image does not take advantage of the full dynamic range.

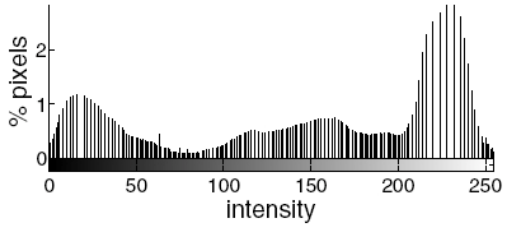


Fig.4(A)Original image histogram

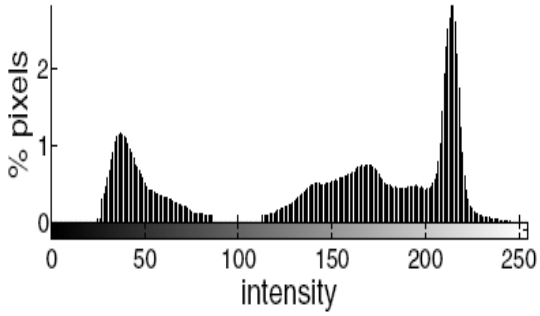


Fig 4(B)SMQT8 enhanced image histogram.

For example -:

In the above fig.4(A) fingerprint image intensities between 0−25, 90−110 and 230−255 in the graph are poorly shown. In the fig4(B)SMQT(SuccessiveMean Quantization Transform) enhanced fingerprint image histogram retains the basic shape of the original but stretches it show the whole dynamic range. Hence, the SMQT adapts the shape of the histogram by performing a nonlinear stretch. The nonlinear properties of the SMQT will yield a balanced stretch of the histogram. These desirable properties come from the nonlinear successive quantization based on the mean

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The histogram equalization has some problems with over saturation and artifacts in several areas area in the images. Notice how the histogram equalized images have a tendency to get washed out or unnatural. These effects do not occur, or are very limited, in the SMQT enhanced images

Conclusion

In this paper we present an effective noise removal approach from the fingerprint images. Here we are using SMQT (Successive Mean Quantization Transform) is a straight forward way to enhance fingerprint image. Here SMQT is use directly on the fingerprint gray scale images. The biometric technology can be successfully applied when the Fingerprint images have good quality.

References:

1.A. K. Jain,Patrick Flynn,Arun A.Ross . "Handbook of Biometrics".
2.D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, Handbook of Fingerprint Recognition. Springer, 2003.
3. Anil K. Jain, *Fundamentals of Digital Image Processing*,Prentice-Hall, 1989, ISBN 0-13-336165
4.. Gray-Scale Image Enhancement using the SMQT, *Mikael Nilsson, Mattias Dahl, and Ingvar Claesson*
5. The Henry Clas sification System Copyright © 2003 International Biometric Group
6. Hong, L., Wan, Y. & Jain, A. (1998) Fingerprint Image Enhancement: Algorithm and Performance Evaluation. IEEE Transactions on Pattern Analysis and Machine Intelligence 20: 777–789
7. Greenberg, S., Aladjem, M., Kogan, D. & Dimitrov, I.(2000) Fingerprint Image Enhancement using FilteringTechniques. 15th International Conference on patternRecognition, Barcelona, vol. III, pp. 326–329.20.
8. Bahuguna, R.D. and Corboline, T.: Prism fingerprint sensor that uses a holographic optical element, *APPLIED OPTICS*,Vol. 35, No. 26 (1996).
9.W. Wang, J.w. Li, F. Huang and H. Feng, " Design and implementationof Log-Gabor filter in fingerprint image enhancement, " PatternRecognition Letters 29 (2008), pp.301–308.Pearson prentice hall, New Jersey, 2004.
10. A. K. Jain and S.Pankanti , "Fingerprint classification and matching, "In A. Bovik, Ed., Handbook for Image and Video Processing . Academic Press, 2000.
11. Z.W. bo, N. X. bao and W.C. jian, " A fingerprint matching algorithm based on relative topological relationship among minutiae, " IEEE Int.Conference Neural Networks & Signal Processing Zhenjiang, China,2008.

Authors Profile

Dr.R.Seshadri was born in Andhra Pradesh, India, in 1959. He received his **B.Tech** degree from Nagarjuna University in 1981. He received his **M.E** degree in Control System Engineering from PSG College of Technology, Coimbatore in 1984. He was awarded with **PhD** from Sri Venkateswara University, Tirupati in 1998. He is currently Director, Computer Center, S.V.University, Tirupati, India. He has Published number of papers in national and international conferences, seminars and journals. At

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

present 12 members are doing research work under his guidance in different areas

Mr.YaswanthKumar .Avulapati received his **MCA** degree with **First class** from Sri Venkateswara University, Tirupati. He received his **M.Tech** Computer Science and Engineering degree with **Distinction** from Acharya Nagarjuna University, Guntur. He is a research scholar in S.V.University Tirupati; Andhra Pradesh.He has presented number of papers in national and international conferences, seminars and journals. He attended number of work shops in different fields.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Conversion of English Text to Braille Code vibration signal for Visually Impaired People

M.Rajasenathipathi
Assistant professor in Computer Science
N.G.M. College of Arts and Science
Pollachi- 642001, Tamilnadu, India
Senathi_pathi@yahoo.co.in

Dr.M.Arthanari
Director
Bharathidasan School of Computer Applications
Ellispettai-638116, Tamilnadu, India
arthanarimsvc@gmail.com

M.Sivakumar
Doctoral Research Scholar
Anna University, Coimbatore, Tamilnadu, India
sivala@gmail.com

***Abstract* - This paper is concerned with the translation of text to Braille code vibration signal in an electronic hand glove using an algorithm based on mathematical logic. It is built on a translation system that combines a finite state machine with left and right context matching and a set of translation rules. This allows the translation of different languages and different grades of Braille contraction, and text-to-Braille conversion. All the implementations perform translation correctly on a range of different operating systems and machines, demonstrating that they are platform-independent. English to Braille code signal vibration conversion responds to increased demands on the Braille code that is integrated education of blind children. The changes inherent in English To Braille code signal conversion are mostly minor for literary Braille and most evident for mathematics and science notations making Braille easier to learn by all stakeholders and easier to read and write for blind people. English to Braille Conversion is optimised for use by students integrated into regular schools, ideal for students using English as their primary or secondary languag; and well-suited to the needs of blind students in developing countries**

Keywords: Braille; cell; hand glove; viberation; characte; dot

I. INTRODUCTION

Braille is a system of writing that uses patterns of raised dots to inscribe characters on paper. It therefore allows visually-impaired people to read and write using touch instead of vision .It is a way for blind people to participate in a literate culture. First developed in the nineteenth century, Braille has become the pre-eminent tactile alphabet. Its characters are six-dot cells, two wide by three tall. Any of the dots may be raised, giving 26 or 64 possible characters. Although Braille cells are used world-wide, the meaning of each of the 64 cells depends on the language that they are being used to depict. Different languages have their own Braille codes, mapping the alphabets, numbers and punctuation symbols to Braille cells in accordance with the varying needs. Braille characters can also be used to represent whole words or groups of letters. This text can then be translated into Braille vibration signal in a hand glove. The glove contains six vibration positions which equals to six raised dots of a Braille cell. Braille translation is not a trivial task, however, because there is a need to perform the contractions correctly.

II. BRAILLE CODE CONVERSION

A. What is Braille code

As mentioned earlier Braille generally consists of cells of six raised dots arranged in a grid of two dots horizontally by three dots vertically. The dots are conventionally numbered 1, 2, and 3 from the top of the left column and 4, 5, and 6 from the top of the right column.The presence or absence of dots gives the coding for the symbol.

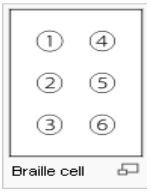


Figure 1 . A Braille Cell

English Braille is used to code the letters punctuation symbols, some double letter signs and word signs directly but capital letters and numbers are dealt with by using a prefix symbol

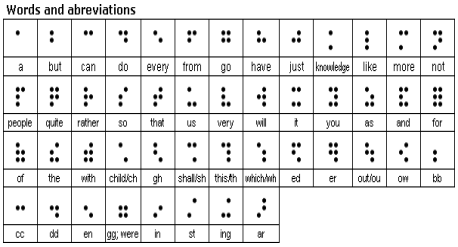


Figure 2. Words and Abbreviations of Braille cell

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Sample texts in Braille

Be kind to others

Figure 3. Example for Braille to English

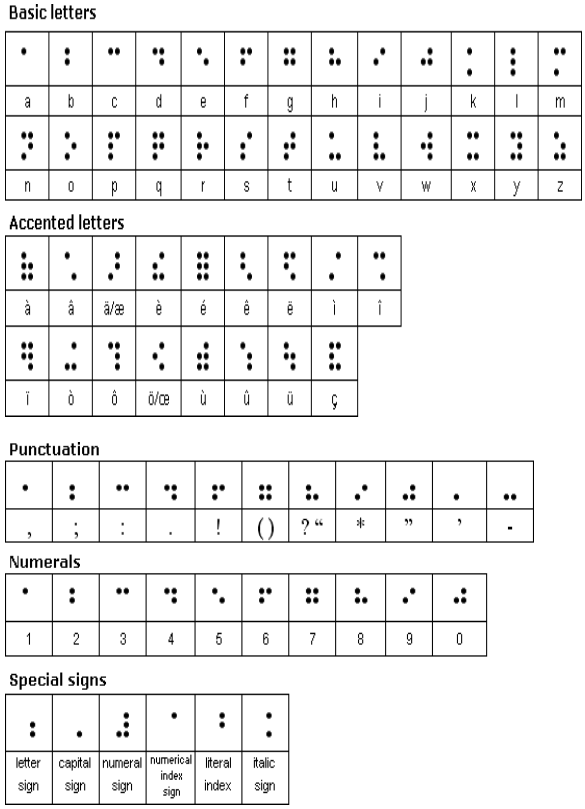


Figure 4. Letters, Numerals and Special symbols of Braille

B. Use of Braille with computer technology

The recent development of personal computers has brought new benefits to Braille users that have access to them. Computerized Braille translation is one of these benefits, but it is not a straightforward process because of the complexities of Braille construction.

- *Difficulties of computer Braille translation:*

Construction makes Braille translation difficult. Without it, Braille translation would be a relatively text characters to Braille cells. Each language (English, French) would need its own unique mapping dictionary, but the operation would be trivial. Contraction greatly increases complexity, so computer translation is generally difficult. For example, in German, there are many words that are formed from the concatenation of other words, similar to the English.

- *Justification of computer Braille translation:*

Braille has come under attack in recent years, with studies showing decreasing usage due to changing patterns of education. Alternative technologies, like speech synthesis, now exist for many of its applications. It can be safely assumed that sighted people will use without complexity, but it is difficult for partially sighted or blind. There are fundamental reasons for continuing with Braille.

- *Reading and writing Braille code is a form of literacy:*

Literacy is a vital component of modern knowledge and society, and Braille code permits literacy for blind people. Using audio only technology denies literacy to Braille users.

- *Braille code is silent:*

Alternative technologies to Braille may not be useful in all circumstances. A speech synthesizer and speech interpreter forbid complete privacy and intrude on the local environment. This would not be appropriate in, for example, an office environment.

- *Braille code is accurate in reading:*

Speech synthesis is an alternative method to Braille code conversion. it may introduce different meanings. Problems will arise from words not known to the synthesizer, miss-spelt words, or with words pronounced differently according to context Reading straight from the text removes a potential source of error.

- *Braille code allows interpretation of the text by the reader, not by another:*

Any reader who wishes to interpret a text themselves – an actor, or a reader for pleasure – may want to form their own interpretation of a text, not be forced into that of a wrong logic in computer program or particular actor. In accuracies in interpretation may be possible.

- *Braille code is cheaper than computer technology***:**

The mechanism to produce Braille vibration signal and produce Blind can understand the communication to the world very easily and it is new communication path for blind persons.

III BRAILLE CODE TRANSLATION MECHANISM

With the profitable commercialization of several public translators the number of approaches available for study in the public realm has decreased. Approaches have inclined to be based around the use of dictionaries of specific translations and limited rule systems because of the trade-off between translation rule numbers and accuracy. The rules work with a ‘window’ of the input text that can potentially be translated, and examine the context to the right of the window to see a rule is correct to determine whether the window is in the middle or end of a word. A finite state machine determines whether the translation is performed. This may be still the basis for commercial products, but they have ceased to be published in the public domain.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A finite state machine, the translation engine, works with, but is independent of a single language rules table. The former contains the translation algorithms and functions, and the latter all of the translation information for translation of one language in one direction. The translation engine can use any language rules table, so any language can be translated to or from Braille code if the language rules table is constructed. The language rules table consists of a set of translation rules and a decision table. During translation the engine works along the input text, character by character. It attempts to match a window of input text starting with the current character with one of the translation rules in the language rules table. A successful match with a translation rule must match a segment of text, the context, to the left and right of the window and the state of the engine. The engine state is controlled by a finite state machine, using the contents of the decision table, and regulates which subset of the language translation rules can be used. This allows both contracted and non-contracted translation can be supported from the same language rules table. The translation rule then provides the translation for that window of input text, which is appended to the growing output text, and the engine moves along the input text to the next unmatched character. A description of the structure of the language rules table will clarify the translation mechanism.

A. The translation algorithm and process

The details of the language rules table provided in the preceding sections should allow the algorithm of the translation engine to be understood. It initializes and loads language translation table and translate input text into Braille code then it send signals to operate Braille hand glove.

Current character = first character of input
While *current character <> end of input*
Do
Begin
Start at first rule whose focus begins with current character
Match = FALSE
Do
If *focus matches*
And *state is ok*
And *right context matches*
And *left context matches*
Begin
Add output from matching rule to output buffer
Set new state according to matching rule
Match = TRUE
Current character moved along input by length of focus
End
Else
Go to next rule
Until *match*
End

Before translation starts, the input text is normalized, ensuring that any character in the text is appropriate for that language. This may be upper and lower case, for instance. It can simplify translation and can also allow more flexibility if required.

The algorithm is high-level, but simple. The design does not make explicit what should output when no match is found for an input character. A space, or the input character, or another character, or nothing may be written, at the discretion of the implementer and produce the no vibration in Braille glove as output. The choice will reflect the final user requirements for an implementation of the system.

IV SYSTEM DESIGN AND IMPLEMENTATION

A. Braille Transalation implementation

Braille Trans is coded in the Matlab 7.6. It is designed to be compiled and run on one of the Microsoft Windows 32-bit operating systems (Windows '95/'98/Me or Windows NT/XP). When compiled to an executable program, in native code, it resides on a machine as a Windows Dynamic Linked Library (DLL). This is a binary executable that supplies a public interface to the Windows operating system and can thus be utilized by other Windows applications with comparative case.

B. Hardware System Description

The block diagram of Braille vibration hand glove system is as follows

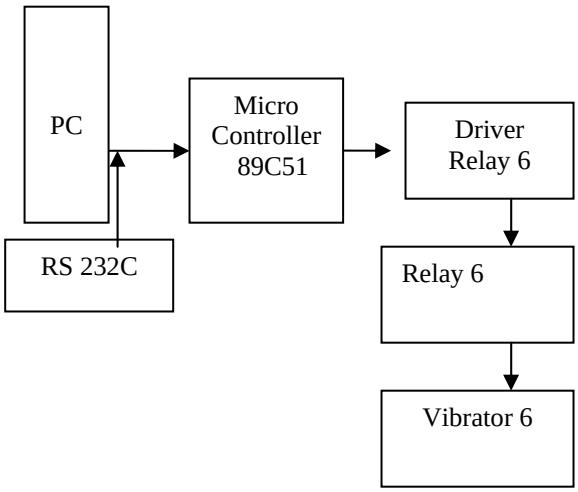


Figure 5. Block diagram for Vibration Hand Glove

The vibration glove system will be comprised of the following key components

1. 89C51 Micro controller
2. Driver Relay
3. Relay
4. RS 232 C
5. Vibrator motor with hand glove

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Figure 6. Braille Vibration Glove

- *Micro controller and Working principle:*

Microcontroller is a general purpose device, which integrates a number of the components of a microprocessor system onto single chip. It has inbuilt CPU, memory and peripherals to make it as a mini computer. A microcontroller is integrated with

1. CPU Core
2. RAM and ROM
3. Some parallel digital i/o ICs

Microcontrollers will combine other devices such as:

A timer module to allow the microcontroller to perform tasks for certain time periods. A serial I/O port to allow data to flow between the controller and other devices such as a PIC or another microcontroller. An ADC allows analogue input data to convert into digital data.

Micro controller is a stand alone unit, which can perform functions on its own without any requirement for additional hardware like I/O ports and external memory. The heart of the microcontroller is the CPU core. In the past, this has traditionally been based on a 8-bit microprocessor unit. For example Motorola uses a basic 6800 microprocessor core in their 6805/6808 microcontroller devices.

In the recent years, microcontrollers have been developed around specifically designed CPU cores, for example the microchip PIC range of microcontrollers.AT89C51 is the 40 pins, 8 bit Microcontroller manufactured by Atmel group. It is the flash type reprogrammable memory. Advantage of this flash memory is we can erase the program within few minutes. It has 4KB on chip ROM and 128 bytes internal RAM and 32 I/O pin as arranged as port 0 to port 3 each has 8 bit bin .Port 0 contains 8 data line(D0-D7) as well as low order address line(AO-A7).

Port 1 contains higher order address line (A8-A15). Port 3 contains special purpose register such as serial input receiver register SBUF, interrupt INT0,INT1 and timers T_0 , T_1 many of the pins have multi functions which can be used as general purpose I/O pins (or) Special purpose function can be decided by the programmer itself.

The AT89C51 is a low-power, high-performance CMOS 8-bit microcomputer with 4KB of Flash Programmable and Erasable Read Only Memory (PEROM). The device is manufactured using Atmel high density non-volatile memory technology and is compatible with the industry standard MCS-51™ instruction set and pin out.

The on-chip Flash allows the program memory to be reprogrammed in-system or by a conventional nonvolatile memory programmer. Here the vibration motor timing and on/off procedure is programmed in hi-tech c language and it is loaded in micro controller. Based on the signals from Braille code value it is activated with the corresponding motors in Braille hand glove.

- *RS-232:*

In telecommunications, RS-232 is a standard for serial binary data interconnection between a *DTE* (Data Terminal Equipment) and a *DCE* (Data Circuit Terminating Equipment). It is commonly used in computer serial ports.

In this circuit the MAX 232 IC used as level logic converter. The MAX232 is a dual driver/receiver that includes a capacitive voltage generator to supply EIA 232 voltage levels from a single 5V supply. Each receiver converts EIA-232 to 5V TTL/CMOS levels. Each driver converts TLL/CMOS input levels into EIA-232 levels. In this circuit the microcontroller transmitter pin is connected in the MAX232 T2IN pin which converts input 5V TTL/CMOS level to RS232 level. Then T2OUT pin is connected to receiver pin of 9 pin D type serial connector which is directly connected to PC. Here the Braille hand glove is connected to COM1 port of PC and it retrieves the Braille signals through RS-232.

- *Vibrator motor:*

The main component in Braille glove is vibration motor. it is configured in two basic varieties ie coin (or flat) and cylinder (or bar). Cylinder type motors are simple brush motors with a traditional axial design. The centric movement of the weight attached to the rotor provides vibration during operation. In Braille glove it is best suited in finger positions. The amount of vibration is directly proportional to the voltage applied to the motor. Cylinder motors are manufactured in high volumes and are fairly inexpensive. An electrical current applied to the coil in the direction of the arrow generates upward force on the left side of the coil and downward force on the right side, causing the coil to revolve clockwise.

V. CONCLUSION AND FURTHER WORK

The development of computer text and Braille translation is necessary for visually impaired community. A number of private translation programs exist, but their private

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

ownership restricts open development. A translation algorithm has been developed and implemented in a Matlab 7.6 translation program and high-tech C language was used in Braille hand glove. However, a number of limitations of this implementation were identified: the platform and character-set dependency of the Matlab implementation, its standalone function, and the difficulties of using Braille Trans as part of a larger application. The project therefore proposed the development of a number of new classes to improve wherever possible on Braille Trans. The development of these classes was a success, and the time available was appropriate to the work required.

REFERENCES

[1] S.G. Basu ,A.lachikawa ," Dialogue Languages and Persons with Disabilities", IEICE transactions , Vol. E87-D, No 6,July 2004, pp 31-43.

[2]Blenkhorn ," A system for converting Braille into print", IEEE Transactions on Rehabilitation Engineering , Vol. 3 no 2, August 1995 pp 215-221 .

[3]H. Choi ,W. Lee W ," Tactile Display as Braille Display for the Visually Disabled ", *in proceedings of IEEE/RSJ International Conference on Intelligent Robotics and Systems*, July 2004, pp 1985-1990 .

[4] J. P. Fritz Barner ," Design of a Hepatic Visualization System for People with Visual Impairments ",IEEE Transactions on Rehabilitation Engineering, vol. 7, No 3, July 1999, pp 372-384 .

[5] Fukumoto et al ," Body coupled Finger Ring wireless wearable keyboard ", *proceeding of the ACM conference on Human factors in computing systems ,France* ,1997 pp 147-154 .

[6] J.M.Gill , priorities for technical research and development for visually disabled persons, *world Blind Union Res. Committee in 1992.*

[7] P.Golding ," Design Study on a Braille Typewriter ", *Western Australia, Curtin University of Technology*. in 1998.

[8] W.Slaby,"Computerized Braille translation", Microcomputer Applications ,Vol.13, June 1990, pp 107-113.

[9] C .Sleng ,Y Lau ,"Regular feature extraction for recognition of Braille". *In Third international conference on Computational Intelligence and Multimedia Applications,ICCIM'99* Proceedings, August 1999 , pp 302-308.

M.Rajasenathipathi (senathi_pathi@yahoo.co.in) is a third year Doctoral Research Scholar in the Research and Development Centre of Bharathiar University. He received his Master degree in computer Applications and M.Phil degree in computer science from M.K. University. He is also working as an Assistant Professor of Computer Science at N.G.M.College , Pollachi , India. He has been an active developer of systems for people with disabilities for the past 10 years.

. Dr.M.Arthanari (arthanarimsvc@gmail.com) holds a Ph.D in Mathematics from Madras University as well as Masters degree in Computer Science from BITS, Pilani. He has worked as a Professor and Head , Department of CSE and IT in Tejaa Shakthi Institute of Technology for Women, Coimbatore, India and Head of the Institution in various colleges in India, He holds a patent issued by the Govt. of India for the invention in the field of Computer Science. He has directed teams of Ph.D researchers and industry experts for developing patentable products. He teaches strategy, project management, creative problem solving, innovation and integrated new product development for last 35 years.

Sivakumar M (sivala@gmail.com) has 10+ years of experience in the software industry including Oracle Corporation. He received his Bachelor degree in Physics and Masters in Computer Applications from the Bharathiar University, India. He holds a patent for the invention in embedded technology. He is technically certified by various professional bodies like ITIL, IBM Rational Clear case Administrator, OCP - Oracle Certified Professional 10g and ISTQB.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

STUDY OF MULTIMEDIA WATERMARKING TECHNIQUES

Mrs. Chhaya S. Gosavi
Department of Computer Engg
Cummins College of Engg. For Women,
Karvenagar,Pune, India
Email: chhaya.gosavi@cumminscollege.in

Dr. C.S. Warnekar
Department of Information Technology
Cummins College of Engg. For Women,
Karvenagar,Pune, India
Email: cswarnekar@hotmail.com

***Abstract*—With the recent burgeoning of networked multimedia systems, techniques are needed to prevent illegal copying / forgery in distributed digital audio/ visual/text document. It may be also desirable to determine where and by how much the multimedia file has been changed from the original due to attacks. This is attributed to increasing instances of hacking during digital communication**

Digital watermarking has been proposed as a solution to the above problem to protect multimedia document. There are two important issues that watermarking algorithms need to address. Firstly, watermarking schemes are required to provide trustworthy evidence for protecting rightful ownership. Secondly, good watermarking schemes should satisfy the requirement of robustness and resist distortions due to common manipulations (such as truncation, compression etc.)

In this paper, various techniques to secure Multimedia data are discussed.

Keywords-Digital watermarking;DCT;IDCT;DFT; DWT; Singular value decomposition; Security.

I. INTRODUCTION

The rapid evolution of the cyber world has greatly facilitated the manipulation and transmission of digital documents in text, images, audio, and video forms. Easy access and replication, however, have led to serious problems regarding copyright protection and/ or distortion prevention of multimedia documents. Conventionally watermarking is used for copyright protection of documents. Presently digital watermarking as an offshoot of computer technology has widened its field of application. Drawing from many related fields, such as cryptography, communication theory, information theory, etc., digital watermarking is proving to be a powerful security measure in transmission of multimedia digital documents. Media owners use this technique to insert identifying information into their document for the purpose of copyright protection. Alternatively they may embed the desired signal into another multimedia document for more secured communication process.

Three qualities are required in digital watermarking: transparency, robustness, and capacity. Transparency refers to the fact that a watermark embedded image signal closely resembles its original version. E.g. it is difficult to differentiate between an audio signal with watermark and its unmarked version. Robustness refers to ability to resist distortion. This is taken care by the invariant properties of the transform. Capacity refers to percentage of watermark signal which may be embedded in original signal without noticeable distortion in the quality. However these characteristics are often mutually contradictory, so compromises must be made while applying them.

Most of the existing watermarking algorithms are applicable to images or video signals. However, the literature on intermixing of audio-visual signals to realize watermarking is comparatively limited. The widespread use of the Internet and the digital audio distribution in MP3 form has made the copyright protection of digital audio work also more and more necessary. Some research works have been published on audio to audio watermarking. These approaches work in the time domain [1], temporal domain [2], DCT domain [3], DWT domain [4], cepstrum domain [5, 6], or sub band domain [7, 8].

In this paper we provide a survey of the latest techniques that are employed to watermark images, audio and video. The paper is organized in the following sections. In Section 2 we describe Image watermarking techniques. In Section 3 we identify the techniques for audio watermarking. In Section 4 we discuss the video watermarking techniques. We conclude this paper in section 5 where we give some guidelines on developing robust watermarking algorithms.

II. IMAGE WATERMARKING

Basically there are two main types of watermarks that can be embedded within an image.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Pseudo-Random Gaussian Sequence

A Gaussian sequence watermark is a sequence of numbers comprising 1 and -1 and which has equal number of 1's and -l's is termed as a watermark. It is termed as a watermark with zero mean and one variation. Such watermarks are used for objective detection using a correlation measure.

Binary Image or Grey Scale Image Watermarks

Some watermarking algorithms embed meaningful data in form of a logo image instead of a pseudo-random Gaussian sequence. Such watermarks are termed as binary image watermarks or grey scale watermarks. Such watermarks are used for subjective detection.

Based on the type of watermark embedded, an appropriate decoder has to be designed to detect the presence of watermark. If it's a pseudo random Gaussian sequence hypothesis, testing is done to detect the presence of watermark. Suppose W is the original watermark bit sequence and W' is the extracted watermark bit sequence, then we can calculate bit error rate (BER) to detect the presence of watermark. If the BER is zero it indicates the presence of watermark; however, if it is one, it indicates absence of watermark. BER is calculated as follows. Suppose D is the retrieved signal and N is the number of bits in watermark then:

$$D = \begin{Bmatrix} 1 \ if \ W_i \neq W_i^{'} \\ 0 \ if \ W_i = W_i^{'} \end{Bmatrix} \quad BER(W,W') = \frac{\sum D}{N} \qquad \ldots\ldots(1)$$

Images can be represented in spatial domain and transform domain. The transform domain image is represented in terms of its frequencies; however, in spatial domain it is represented by pixels. In simple terms transform domain means the image is segmented into multiple frequency bands. To transfer an image to its frequency representation we can use several reversible transform like Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), or Discrete Fourier Transform (DFT). Each of these transforms has its own characteristics and represents the image in different ways. Watermarks can be embedded within images by modifying these values, i.e. the pixel values or the transform domain coefficients. Simple watermarks could be embedded in the spatial domain of images by modifying the pixel values or the least significant bit (LSB) values; however, more robust watermarks could be embedded in the transform domain of images by modifying the transform domain coefficients. Following Figures shows the result of Spatial domain technique, i.e. LSB modification.

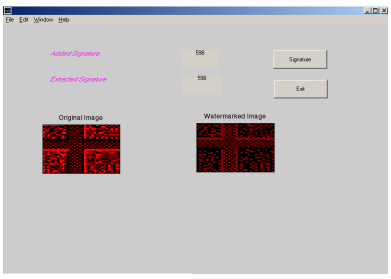


Figure 1. Embedding Watermark (LSB)

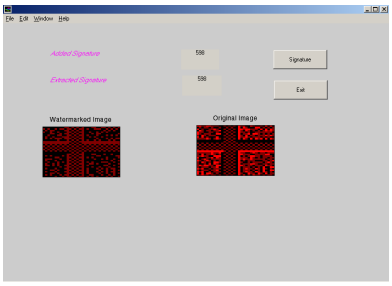


Figure 2. Extracting Watermark (LSB)

A. DCT DOMAIN WATERMARKING

DCT based watermarking techniques are more robust compared to simple spatial domain watermarking techniques. Such algorithms are robust against simple image processing operations like low pass filtering, brightness and contrast adjustment, blurring etc. However, they are difficult to implement and are computationally more expensive. At the same time they are weak against geometric attacks like rotation, scaling, cropping etc. DCT domain watermarking can be classified into **Global DCT** watermarking and **Block based DCT** watermarking. One of the first algorithms presented by Cox et al. (1997) used global DCT approach to embed a robust watermark in the perceptually significant portion of the Human Visual System (HVS). Embedding in the perceptually significant portion of the image has its own advantages because most compression schemes remove the perceptually insignificant portion of the image. In spatial domain it represents the LSB however in the frequency domain it represents the high frequency components. The main steps of any block based DCT algorithm are as follows:

Steps in DCT Block Based Watermarking Algorithm

1) Segment the image into non-overlapping blocks of 8x8

2) Apply forward DCT to each of these blocks

3) Apply some block selection criteria (e.g. HVS)

4) Apply coefficient selection criteria (e.g. highest)

5) Embed watermark by modifying the selected coefficients.

6) Apply inverse DCT transform on each block

Most algorithms are classified based on step 3 and 4 i.e. the main difference between most algorithms is that they differ either in the block selection criteria or coefficient selection criteria. Based on the perceptual modeling strategy incorporated by the watermarking algorithms they could be classified as algorithms with:

1) No Perceptual Modeling:

Such algorithms do not incorporate any perceptual modeling strategy while embedding a watermark.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

2) Implicit Perceptual Modeling

Such algorithms incorporate the transform domain properties for perceptual modeling. The coefficient selection criterion is as follows:

i) Select those transform coefficients which have large perceptual capacity, because they allow stronger watermarks to be embedded and result in least perceptual distortion. DC component satisfy this criteria and hence can be used.

ii) Select only those coefficients which are least changed by common image processing attacks like low-pass filtering, noise addition etc. Low frequency AC components (or high magnitude coefficients) as well as high magnitude DC components satisfy the above criteria and can be selected.

iii) High frequency components are affected by common image processing operations hence they are not a good choice for watermarking.

3) Explicit Perceptual Modeling

Such algorithms incorporate the HVS properties for perceptual modeling. HVS models allow us to raise or lower the strength of the watermark because it takes into account the local image properties like contrast, brightness, variance etc.

B. DWT DOMAIN WATERMARKING

In the last few years wavelet transform has been widely studied in signal processing in general and image compression in particular. In some applications wavelet based watermarking schemes outperforms DCT based approaches.

Characteristics of DWT

i) The wavelet transform decomposes the image into three spatial directions, i.e. horizontal, vertical and diagonal. Hence wavelets reflect the anisotropic properties of HVS more precisely.

ii) Wavelet Transform is computationally efficient and can be implemented by using simple filter convolution.

iii) Magnitude of DWT coefficients is larger in the lowest bands (LL) at each level of decomposition and is smaller for other bands (HH, LH, HL).

iv) The larger the magnitude of the wavelet coefficient the more significant it is.

v) Watermark detection at lower resolutions is computationally effective because at every successive resolution level there are few frequency bands involved.

vi) High resolution subbands helps to easily locate edge and textures patterns in an image.

Advantages of DWT over DCT

i) Wavelet transform understands the HVS more closely than the DCT.

ii) Wavelet coded image is a multi-resolution description of image. Hence an image can be shown at different levels of resolution and can be sequentially processed from low resolution to high resolution.

iii) Visual artifacts introduced by wavelet coded images are less evident compared to DCT because wavelet transform doesn't decompose the image into blocks for processing. At high compression ratios blocking artifacts are noticeable in DCT; however, in wavelet coded images it is much clearer.

iv) DFT and DCT are full frame transform, and hence any change in the transform coefficients affects the entire image except if DCT is implemented using a block based approach. However DWT has spatial frequency locality, which means if signal is embedded it will affect the image locally. Hence a wavelet transform provides both frequency and spatial description for an image.

Disadvantages of DWT over DCT

Computational complexity of DWT is more compared to DCT. As Feig (1990) pointed out it only takes 54 multiplications to compute DCT for a block of 8x8, unlike wavelet calculation depends upon the length of the filter used, which is at least 1 multiplication per coefficient.

DWT Watermarking

DWT based watermarking schemes follow the same guidelines as DCT based schemes, i.e. the underlying concept is the same; however, the process to transform the image into its transform domain varies and hence the resulting coefficients are different. Wavelet transforms use wavelet filters to transform the image. There are many available filters, although the most commonly used filters for watermarking are Haar Wavelet Filter, Daubechies Orthogonal Filters and Daubechies Bi-Orthogonal Filters. Each of these filters decomposes the image into several frequencies. Single level decomposition gives four frequency representations of the images. These four representations are called the LL, LH, HL, HH subbands as shown in Fig.3.

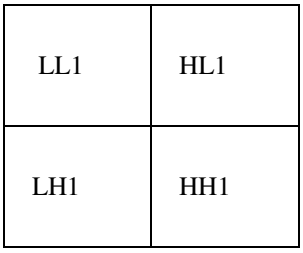


Figure 3. *Single level Decomposition using DWT*

DWT algorithms can be classified based on their decoder requirements as Blind Detection or Non-blind Detection. Blind detection doesn't require the original image for detecting the watermarks; however, non-blind detection requires the original image.

C. DFT DOMAIN WATERMARKING

DFT domain has been explored by researches because it offers robustness against geometric attacks like rotation, scaling, cropping, translation etc.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Characteristics of DFT

i) DFT of a real image is generally complex valued, which results in the phase and magnitude representation of an image.

ii) DFT shows translation invariance. Spatial shifts in the image affects the phase representation of the image but not the magnitude representation, or circular shifts in the spatial domain don't affect the magnitude of the Fourier transforms.

iii) DFT is also resistant to cropping because effect of cropping leads to the blurring of spectrum. If the watermarks are embedded in the magnitude, which are normalized coordinates, there is no need of any synchronization

iv) The strongest components of the DFT are the central components which contain the low frequencies.

v) Scaling of image results in amplification of extracted signal and can be detected by correlation coefficient. Translation of image has no result on extracted signal.

vi) Rotation of image results in cyclic shifts of extracted signal and can be detected by exhaustive search.

vii) Scaling in the spatial domain causes inverse scaling in the frequency domain. Rotation in the spatial domain causes the same rotation in the frequency domain.

Co-efficient Selection Criteria

i) Modification to the low frequency coefficients can cause visible artifacts in the spatial domain. Hence, low frequency coefficients should be avoided

ii) High frequency coefficients are not suitable because they are removed during JPEG compression.

iii) The best location to embed the watermark is the mid frequency.

Advantages of DFT over DWT and DCT

DFT is rotation, scaling and translation (RST) invariant. Hence it can be used to recover from geometric distortions, whereas the spatial domain, DCT and the DWT are not RST invariant and hence it is difficult to overcome from geometric distortions.

There are two different kinds of DFT based watermark embedding techniques. One in which watermark is directly embedded or template based embedding.

D. FFT AND DHT DOMAIN WATERMARKING

FFT is robust against compression and RST attacks. It is a template based embedding algorithm. Apart from the template, an informative watermark is embedded to prove ownership. In case the image undergoes a geometric distortion the template is reversed back to its original location and then the watermark is extracted.

The DHT based watermarking techniques rely on the Discrete Hadamard Transform. Initially the multi-resolution Hadamard transform is applied to the image to decompose it into various frequency bands like low-low, low-high and high-high. The lowest frequency band is then divided into 8x8 blocks and 2D complex Hadamard transform is applied. Watermark is embedded in this domain by altering the phase component of the most significant image component. The watermark is embedded in the phase component because phase modulation is more robust to noise than amplitude modulation.

As stated above watermarking schemes can be applied in the time domain or frequency domain representation of signal. In all frequency domain watermarking schemes, there is a conflict between robustness and transparency. If the watermark is embedded in perceptually most significant components, the scheme would be robust to withstand attacks but the watermark may be difficult to hide. On the other hand, if the watermark is embedded in perceptually insignificant components, it would be easier to hide the watermark but the scheme may be less resilient to distortions due to attack.

A few years ago, Singular Value Decomposition (SVD) transform was applied to digital watermarking. It may be noted that the mathematical theory of SVD for square matrices was discovered independently by Beltrami in 1873 and Jordan in 1874, and extended to rectangular matrices by Eckart and Young in the 1930s. Later Gene Golub demonstrated its feasibility and usefulness as a tool in a variety of applications. SVD has proved to be one of the most powerful tools of linear algebra. Following figures shows the results obtained by applying SVD for image watermarking.

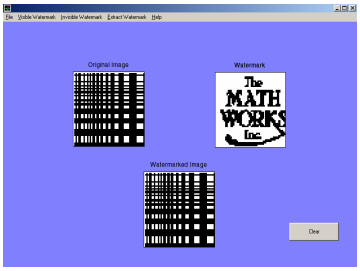


Figure 4. Watermark embedding (SVD)

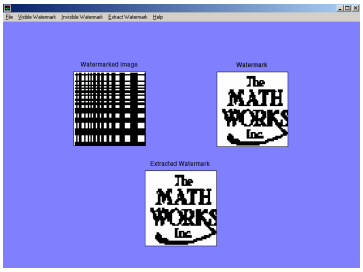


Figure 5. Watermark extraction (SVD)

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

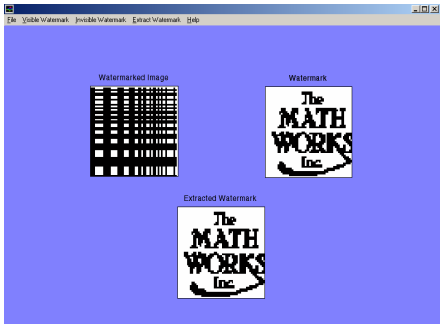


Figure 6. Extracting Watermark from Rotated Image

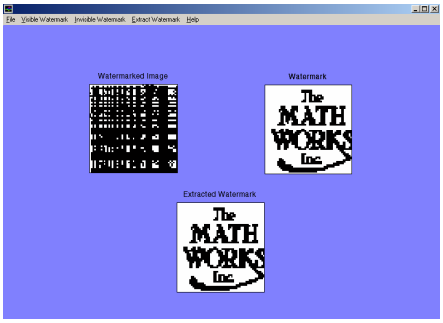


Figure 7. Extracting Watermark from Noised Watermarked Image

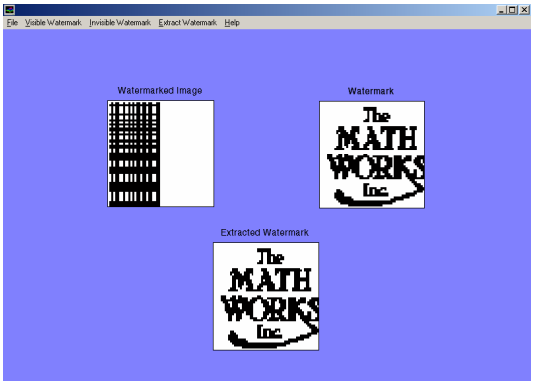


Figure 8. Extracting Watermark from Cropped Watermarked Image

III. AUDIO WATERMARKING

Digital audio watermarking is a technique for embedding additional data along with audio signal. Audio watermarking is a difficult process because of the sensitivity of Human Auditory System (HAS). A number of audio watermarking techniques are exploit different ways in order to embed a robust watermark and to maintain the original audio signal fidelity. These techniques are classified according to the domain where the watermark is embedded to four categories

A. *Frequency Domain Audio Watermarking*

Audio watermarking techniques, that works in frequency domain, take the advantage of audio masking characteristics of HAS to embed an inaudible watermark signal in digital audio. Transforming audio signal from time domain to frequency domain enables watermarking system to embed the watermark into perceptually significant components. This will provide the system with a high level of robustness, because of that any attempt to remove the watermark will result in introducing a serious distortion in original audio signal fidelity. The input signal is first transformed to frequency domain where the watermark is embedded, the resulting signal then goes through inverse frequency transform to get the watermarked signal as output as shown in Figure 9.

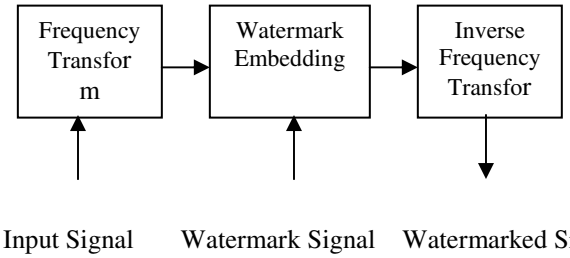


Figure 9. Watermarking in frequency domain

In spread spectrum communication, one transmits a narrowband signal over a much larger bandwidth such that the signal energy present in any single frequency is imperceptible. Similarly the watermark is spread over very many frequency components so that the energy of any component is very small and certainly undetectable. In this method the frequency domain of cover signal is viewed as a communication channel and the watermark is viewed as a signal that is transmitted through it. Attacks and unintentional signal distortions are thus treated as noise that the transmitted signal must be immune to. In order for the watermark to be robust, watermark must be placed in perceptually significant regions of the cover signal despite the risk of potential fidelity distortion. Conversely if the watermark is placed in perceptually insignificant regions, it is easily removed, either intentionally or unintentionally by, for example, signals compression techniques that implicitly recognize that perceptually weak components of a signal need not be represented.

Another transform is cepstrum domain. Cepstrum domain has been widely adopted for phonetic analysis and recognition, which include a series of operations: (1) Fourier transform, (2) take logarithm, and (3) inverse Fourier transform. It is obvious that these three operations are linear and that the original signal in the time domain can be exactly recovered from its cepstrum domain representation. After a general attack, the statistical mean of the cepstrum coefficients for an audio signal

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

experience much less variance. Due to the attack-invariant feature, the watermark information can be preserved.

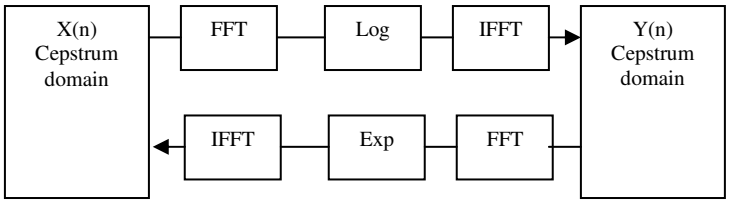
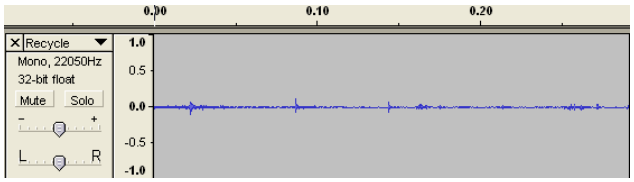


Figure 10. Cepstrum Analysis

We implemented the audio watermarking in cepstrum domain; watermark used is a binary logo image. The sampling rate, fs, was used for playback. The value typically supported by sound cards is 44100 Hz. Each frame had 1024 samples. Each song had duration of 300 seconds and was recorded in mono at a sampling rate of 16 bits. The audio editing and attacking tools adopted in this experiment were Audacity and CoolEdit Pro 2.0 Even after embedding logo image into the audio signal, it has been observed that watermarked audio has very equally good perceptual quality. Using the above extraction algorithm, logo image was then successfully extracted from watermarked audio. This algorithm is also tested for various synchronous attacks like, echo, compress, cut. However results of comparisons with other robust techniques are awaited.

Following figures shows some of the experimental results.

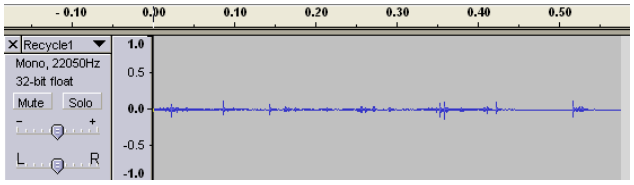
Original Audio:



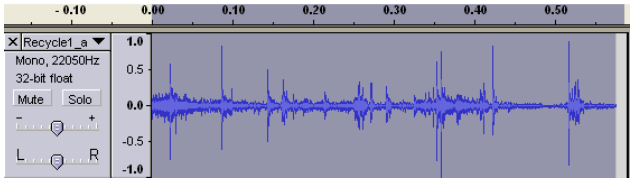
Watermark image

Hi.bmp

After Embedding watermark



After Amplifying



Extracted watermark

Figure 10. Experimental Results-Cepstrum Domain

B. Time Domain Audio Watermarking

In time domain watermarking techniques, watermark is directly embedded into audio signal. No domain transform is required in this process. Watermark signal is shaped before embedding operation to ensure its inaudibility. The available time domain watermarking techniques insert the watermark into audio signal by simply adding the watermark to the signal. Embedding a watermark into time domain involves challenges related to fidelity and robustness. Shaping the watermark before embedding enables the system to maintain the original audio signal fidelity and renders the watermark inaudible. As for robustness, time domain watermarking systems use different techniques to improve the robustness of the watermark technique of digital signals is well known and developed over years.

There are two methods for audio watermarking in time domain. In the first method the watermark signal is modulated using the original audio signal and filtered by lowpass filter to reduce the distortion that might be result from embedding the watermark. The original audio signal is divided into segments and then each segment is watermarked separately by embedding the same watermark. Another watermarking system uses the HAS masking effects to shape the watermark signal. Shaping operation is performed in frequency domain, but the shaped watermark is embedded into audio signal in time domain. Watermark is a noise-like sequence generated by using two keys x1 and x2. The first key x1 is author dependent. The second key x2 is computed from audio signal that the author wants to watermark. It is computed from the signal using a one-way hash function. The two keys are mapped to pseudorandom number generator to generate a noise-like sequence, watermark. Original audio signal is required in detection process to compute the second key x2, and to extract the embedded watermark.

C Compressed Domain Audio Watermarking

A number of techniques are proposed to embed a watermark signal into MPEG audio bit stream, rather than going through decoding/encoding process in order to apply watermarking scheme in uncompressed domain Such systems are suitable for “pay audio” scenario, where the provider stores audio contents in compressed format. During download of music, the customer identifies himself/herself with his/her

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

unique customer ID, which therefore is known to the provider during delivery. In order to embed the customer ID into the audio data using a watermarking technique, a scheme is needed that is capable of watermarking compressed audio on the fly during download. MPEG audio compression is a lossy algorithm and uses the special nature of the HAS. It removes the perceptually irrelevant parts of the audio and makes the audio signal distortion inaudible to the human ear. MPEG encoding process has the following steps:

i. Input audio samples pass through a mapping filter bank to divide the audio data into subbands (subsamples) of frequency.

ii. At the same time, the input audio samples pass through MPEG psychoacoustics model, which creates a masking threshold of audio signal. Masking threshold is used by quantization and coding step to determine how to allocate bits to minimize the quantization noise audibility.

iii. Finally, the quantized subband samples are packed into frames (coded stream).

The MPEG audio stream consists of frames. Frame is the smallest unit which can be decoded individually. Each frame contains audio data, header, CRC (Cyclic Redundancy Code), and ancillary data. In frame, each subband has three groups of samples with 12 samples per group. The encoder can use a different scale factor for each group. Scale factor is determined upon masking threshold and used in reconstruction of audio signal. The decoder multiplies the quantizer output to reconstruct the quantized subband sample. Figure 10 depicts the general format of MPEG frame.

Header	CRC	Bit Allocation	Scale Factor	Encoded Sample	Ancillary Data

Figure. 10 *Frame Format of MPEG Audio*

MPEG audio decoding process is simple a reverse of the encoding process. The decoding takes the encoded bit stream as an input, unpacks the frames, reconstructs the frequency samples (subbands samples) using scale factors, and then inverses the mapping to re-create the audio signal samples.

However, watermarking systems have a number of differences. These differences can be considered in evaluating performance of watermarking systems and suitability of these systems for a specific application. These differences can be explained as follows:

i. Some audio watermarking systems require the original audio signal, or any information derived from it, to be presented in detection process. This will leads to a large number of original works have to be stored and searched during detection. Systems that require the original audio signal are not suitable for some type of applications, in case that detection process has no access to the original work or it is not acceptable to disclose it. On the other hand, presenting the original signal yields in efficient watermark extraction consequently efficient detection. Audio watermarking systems that are based on patchwork algorithm use a statistical detection process (hypothesis testing) and don't need the original audio for detection purpose. In spite of that a number of audio watermarking techniques require only the watermarked signal in detection watermark key is needed in both embedding and detection.

ii. In order to maintain the watermark security, watermark would be embedded into selected regions of some domain transform of audio signal. These regions are selected randomly by generating a sequence of indexes. Sequence generation is paramerized by a key called watermarking key. This key is required in both embedding and detection. In some watermarking systems, watermarking key is used to generate the watermark itself. In this case, the watermark would be a random sequence of bits or digits generated by some sort of algorithms ensure non-invertiblitiy of watermark in order to maintain the security of watermarking key. Watermarking key could be provided by the copyright owner or a combination of information provided by him/her and information derived from original signal. In such case, original signal will be required in detection process for key generation purpose. In all scenarios, the key is used as a seed for random number generator. Sometimes, disclosing the watermarking key or having an access to it becomes impossible. Thus, using the same key in detection and embedding will not be acceptable. A solution to such problem could be found in using two keys, one for embedding and another for detection

iii. During embedding process, original audio signal is divided into frames. Then after, each frame is watermarked separately. Some watermarking systems embed the same watermark into a number of frames to enhance watermark robustness. But, in other systems each frame is watermarked with different watermark.

iv. Because of sensitivity of HAS, watermark signal must be shaped to rent it inaudible. Masking characteristics of audio signal can be used for this purpose. Psychoacoustics MPEG model is commonly used to calculate masking threshold that is used in weighting the watermark.

A general work frame for digital audio watermarking systems can be concluded as follows:

i. Watermarking system should be able to embed any set of data in to audio signal, and the detector should be able to retrieve the embedded data (i.e. not just report that watermark is presented or not)

ii. Watermark embedded (detection) module should be independent of mode of operating. (e.g. the same watermark is embedded into multiple frames of audio signal or different watermark is embedded into each frame).

iii. Watermarking key generation should be independent of watermark embedding and detection (e.g. embedding and detection will not be effected whether original signal is involved in key generation or not).

The above points enable audio watermarking system to be suitable for variety of application and make it possible to put standards and evaluation benchmark

IV. VIDEO WATERMARKING

Most of the video watermarking schemes are based on the techniques of image watermarking and directly applied to raw

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

video or compressed video. However, current image watermarking schemes are not capable of adequately protecting video data. Video watermarking introduces some issues which is not present in image watermarking. Due to large amounts of data and inherent redundancy between frames, video signals are highly susceptible to pirate attacks, including frame averaging, frame dropping, frame swapping, statistical analysis, etc. Applying a fixed image watermark to each frame in the video leads to problems of maintaining statistical and perceptual invisibility. Furthermore, such an approach is necessarily video independent; as the watermark is fixed while the frame changes. Applying independent watermarks to each frame also presents a problem. Regions in each video frame with little or no motion remain the same frame after frame. Motionless regions may be statistically compared or averaged to remove independent watermarks. In addition, video watermarking schemes must not use the original video during watermark detection as the video usually is in very large size and it is inconvenient to store it twice.

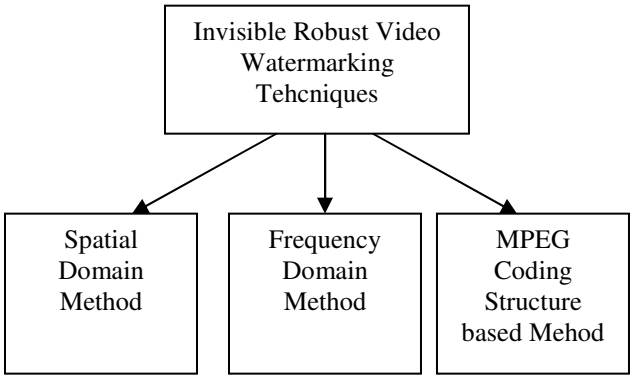


Figure 11. Classification map of existing digital video watermark techniques

A. Spatial Domain Watermarks

Spatial domain algorithms generally share the following characteristics:

i) The watermark is applied in the pixel or coordinate domain.

ii) No transforms are applied to the host signal during watermark embedding.

iii) The watermark is derived from the message data via spread spectrum modulation.

iv) Combination with the host signal is based on simple operations, in the pixel domain.

v) The watermark can be detected by correlating the expected pattern with the received signal.

The main strengths of pixel domain methods are that they are conceptually simple and have very low computational complexities. As a result they have proven to be most attractive for video watermarking applications where real-time performance is a primary concern. However, they also exhibit some major limitations: The need for absolute spatial synchronization leads to high susceptibility to de-synchronization attacks; lack of consideration of the temporal axis results in vulnerability to video processing and multiple frame collusion; and watermark optimization is difficult using only spatial analysis techniques.

The simplest method is to just flip the lowest-order bit of chosen pixels in a grey scale or colour image. This will work well only if the image is subjected to any human or noisy modification. A more robust watermark can be embedded in an image in the same way that a watermark is added to paper. Such techniques may superimpose a watermark symbol over an area of the picture and then add some fixed intensity value for the watermark to the varied pixel values of the image. The resulting watermark may be visible or invisible depending upon the value (large or small, respectively) of the watermark intensity. One disadvantage of spatial domain watermarks is that picture cropping, which is a common operation of image editors, can be used to eliminate the watermark. Spatial watermarking can also be applied using colour separation.

In this way, the watermark appears in only one of the colour bands. This renders the watermark visibly subtle so that it is difficult to detect under regular viewing. However, the watermark appears immediately when the colours are separated for printing or xerography. This renders the document useless to the printer unless the watermark can be removed from the colour band. This approach is used commercially for journalists to inspect digital pictures from a photo-stockhouse before buying non-watermarked versions.

B. Frequency Domain Watermarks

Generally DCT, FFT and wavelet transform are used as the methods of data transformation as seen in section 2 and 3. The main strength offered by transform domain techniques is that they can take advantage of special properties of alternate domains to address the limitations of pixel-based methods or to support additional features. For instance, designing a watermarking scheme in the Discrete Cosine Transform (DCT) domain leads to better implementation compatibility with popular video coding algorithms such as Moving Pictures Experts group (MPEG)-2, and in the shift and rotation-invariant Fourier domains facilitates the design of watermarks that inherit these attractive properties. Besides, analysis of the host signal in a frequency domain is a prerequisite for applying more advanced masking properties of the HVS to enhance watermark robustness and imperceptibility. Generally, the main drawback of transform domain methods is their higher computational requirement.

C. Watermarks Based on MPEG Coding Structures

Video watermarking techniques that use MPEG-1, -2 and -4 coding structures as primitive components are primarily motivated by the goal of integrating watermarking and compression to reduce overall real-time video processing complexity. Compression in block-based schemes like MPEG-2 is achieved by using forward and bi-directional motion prediction to remove temporal redundancy, and statistical methods to remove spatial redundancy. One of the major drawbacks of schemes based on MPEG coding structures is that they can be highly susceptible to re-compression with different parameters, as well as conversion to formats other than MPEG.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Comparison between Different Watermarking Schemes

In general, watermarking schemes can be roughly divided into two categories: spatial domain watermark, and transformed domain watermark. Watermarking schemes in spatial domain are less robust than those in frequency domain. LSB, threshold-based correlation and m-sequence watermarks are perform worse than the other watermarking algorithms. Therefore, these watermarking algorithms are classified as fragile or semi-fragile watermarking. They can be applied for the purpose of proving the integrity of a document. The frequency domain watermarking schemes are relatively more robust than the spatial domain watermarking schemes, particularly in lossy compression, noise addition, pixel removal, rescaling, rotation and shearing. DCT-based watermarking scheme is the most robust to lossy compression. Moreover, DWT-based watermarking scheme is the most robust to noise addition. DFT-based watermarking scheme with template matching can resist a number of attacks, including pixel removal, rotation and shearing. The purpose of the template is to enable resynchronization of the watermark payload spreading sequence. It is a key dependent pattern of peaks, which is also embedded into DFT magnitude representation of the frame. The peaks are not embedded by addition, but rather by modifying the value of the target coefficient, such that it is at least two standard deviations above the mean. Radon transformation resists attacks by resealing and geometric distortion.

V. CONCLUSION

This paper is an attempt to summarize various watermarking techniques used to secure Multimedia. Some of these techniques are implemented and tested for images and audio watermarking. We are still working on video watermarking.

REFERENCES

[1] P. Bassia, I. Pitas, and N. Nikolaidis, "Robust audio watermarking in the time domain," *IEEE Transactions on Multimedia, Vol. 3, 2001,* pp. 232-241.

[2] A. N. Lemma, J. Aprea, W. Oomen, and L. van de Kerkhof, "A temporal domain audio watermarking technique," *IEEE Transactions on Signal Processing, Vol. 51, 2003*, pp. 1088-1097.

[3] I. K. Yeo and H. J. Kim, "Modified patchwork algorithm: a novel audio watermarking scheme," *IEEE Transactions on Speech and Audio Processing, Vol. 11, 2003*, pp. 381-386.

[4] L. Cui, S. X. Wang, and T. F. Sun, "The application of wavelet analysis and audio compression technology in digital audio watermarking," in *Proceedings of the IEEE International Conference on Neural Networks and Signal Processing, Vol. 2, 2003,* pp. 1533-1537.

[5] S. K. Lee and Y. S. Ho, "Digital audio watermarking in the cepstrum domain," *IEEE Transactions on Consumer Electronics, Vol. 46, 2000, pp. 744-750.*

[6] X. Li and H. H. Yu, "Transparent and robust audio data hiding in cepstrum domain," in *Proceedings of the IEEE International Conference on Multimedia and Expo, Vol.1, 2000*, pp. 397-400.

[7] X. Li and H. H. Yu, "Transparent and robust audio data hiding in subband domain," in *Proceedings of the IEEE International Conference on Information Technology: Coding and Computing, 2000,* pp. 74-79.

[8] J. M. Huang, "Key-based audio watermarking system using wavelet packet decomposition," Master's thesis, Dept. of Electrical Engineering, National Central University, Taiwan, 2002.

[9] Shi-Cheng Liu and Shinfeng D. Lin, "BCH Code-Based Robust Audio Watermarking in the Cepstrum Domain," *Journal of Information Science and Engineering 22*, 535-543 (2006)

[10] Ruizhen Liu, Tieniu Tan, "A Svd-Based Watermarking Scheme For Protecting Rightful Ownership", IEEE Transaction on Multimedia, Vol. 4, Issue 1, pp. 121 – 128, March -2002

[11] Mohd Shahidan Abdullah, Azizah Abd Manaf. : An Overview of Video Watermarking Techniques. University Technology of Malaysia, Postgraduate Annual Research Seminar (2007)

[12] D. Mitchell, S.B. Zhu. :Multi resolution Scene-Based Video Watermarking using Perceptual Models. In: IEEE Journal on Selected Areas in Communications, vol. 16, No. 4, (1998)

[13] A. Alattar, M. Celik et. E.Lin. : Watermarking Low Bit Rate advance simple profile MPEG-4 bitstreams. In: Proceeding of the International conference on Acoustics, Speech and Signal Processing, HongKong. (2003)

[14] J. Bloom, I. Cox, T. Kalker, J.-P. Linnartz, M. Miller, C. Traw. : Copy protection for DVD video. In: Proc. IEEE 87 (7) 1267–1276, (1999)

[15] F. Bartolini, A. Manetti, A. Piva, M. Barni. : A data hiding approach for correcting errors in H.263 video transmitted over a noisy channel. In: Proceedings of the IEEE Fourth Workshop on Multimedia Signal Processing, pp. 65–70. (2001)

[16] D. Boneh, J. Shaw. : Collusion-secure fingerprinting for digital data. In : IEEE Trans. Inform. Theory. 44 (5) 1897–1905. (1998)

[17] I. Brown, C. Perkins, J. Crowcroft. : Watercasting: distributed watermarking of multicast media. In: Proceedings of the First International Workshop on Networked Group Communication, Lecture Notes in Computer Science, Vol. 1736, Springer, Berlin, pp. 286–300. (1999)

[18] Ingemar J. Cox, Senior Member, IEEE, Joe Kilian, F. Thomson Leighton, and Talal Shamoon, Member, IEEE. : Secure Spread Spectrum Watermarking for Multimedia. In : IEEE Transactions On Image Processing, Vol. 6, No. 12. (1997)

[19] Nasir Memon, Ping Wah Wong,"Protecting Digital Media Content", Communication Of ACM, Vol. 41,No. 7,pp. 35-40, July 1998

[20] John M. Acken, "How Watermarking Adds Value to Digital Content" Communication Of ACM,Vol. 41,No. 7,pp. 74- 77, July 1998

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Adaptive e-Learning: A Conceptual Solution for the analysis of link between Medium of Instruction and Performance

Damodaran Venkatesan
Director-CUIC, Anna University Tiruchirappalli, Tiruchirapalli,
Tamil Nadu, India
Email : danvenkatesan@yahoo.com

RM.Chandrasekaran
Registrar, Anna University Tiruchirappalli, Tiruchirappalli
Tamil Nadu, India
Email: aurmc@hotmail.com

***Abstract*- e-Learning system emphasizes on the available content to all learners irrespective of their knowledge level and relevance. This paper concentrates on the result oriented investigation of the impact of learner's primary medium of instruction (MI) at school in understanding the content given to them in English and proposes a design for developing an Adaptive e-Learning (AeL) system personalized to the learner. An appropriate statistical model has been developed to analyze the impact and the results implied that the MI (Regional language / English) at school has greater impact on the performance compared to the region (Rural / Urban) the students hailed from, when the same content is given to them in English using e-Learning. Based upon the significant gap in the performance between the rural regional language medium students and urban English medium students, the decision procedure rules for providing personalized content to the learner were developed. The AeL system personalized to the leaner was successfully implemented with the help of an extensive survey to bridge the gap in performance between the regional language medium students and English medium students.**

Keywords-Components: **e-Learning (eL), Adaptive e-Learning (AeL), MI, Multiple Regression, Urban / Rural Capacity gap**

I. INTRODUCTION

Adaptivity is an important issue in e-Learning but very few learning systems incorporate adaptivity in today's e-education. E-Learning platforms such as Moodle and WebCT provide lot of simple features to administer and create courses. As such, they have become very successful in e-education. Adaptive e-Learning (AeL) system creates a personalized environment for the learner. Personalization helps to display required content in understandable and desirable form and prevents unwanted or unrelated content being displayed, thus saving cost and time. The focus of this research is primarily on countries offering school education both in English medium and in their regional languages. In countries like India, English is the MI primarily at urban places, and the regional languages are the same will be primarily at rural places. This study is also applicable to English-speaking countries with student immigrants from across the globe.

It is generally believed that the urban / rural exposure to the student has more impact than the primary MI at school, while understanding the content given to them in English. To test this study, two experiments were performed with a heterogeneous group of first year engineering students both from urban and rural backgrounds and with a combination of regional language (Tamil) and English as their primary MI at school, but with similar pre-requisite subject knowledge.

Organization of the paper

Section 2 deals with an experiment offering an English language test using e-Learning to the students to evaluate their understanding of English. Section 3 deals with an experiment offering 50 minutes of technical content and subject test in English using eL to evaluate the impact of foreign language knowledge to understand a technical subject. A Statistical Regression Mode (SRM) is used to validate the above hypothesis. Subsequently, an AeL system using Moodle with decision procedure rules for providing personalized content to the learner was developed and successfully implemented to bridge the gap in performance between the regional language (Tamil) students and English medium students.

II. IMPACT OF MI IN UNDERSTANDING ENGLISH

The hypothesis of this experiment is that the urban / rural exposure to the student has greater impact than the primary MI at school in understanding the content given to them in English.

A. Participants

A heterogeneous group of 219 number of undergraduate first year engineering students both from urban and rural backgrounds (116 urban and 103 students hailing from rural areas) and with a combination of regional language and English as their primary MI at

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

school (148 regional language and 71 English medium students), but with similar pre-requisite subject knowledge (all of them scored entry level cut-off marks between 170 to 190 marks out of 200 marks to get admitted into the engineering degree programme). Even students, who did their schooling in regional language,

TABLE 1

Medium of Study	Residence	No. of Students	English Test Score - Avg %	Range %
Regional	Rural	90	46.39%	30% - 65%
Regional	Urban	58	51.29%	30% - 65%
English	Rural	13	61.92%	40% - 75%
English	Urban	58	70.52%	50% - 90%

Table 1: English Test Performance

had English as one of their subjects throughout their schooling. The other factors considered were the economical and educational background of their families and the student is a boy or a girl.

B. Evaluation Procedure

The participants were observed in the computer laboratory and given two reading comprehensions in English through eL engine which shows uniform content to all the users. They were given a total of 50 minutes to go through the contents and answer 20 questions based on the contents.

C. Results and decisions

Table 1 summarizes the task performance of the impact of MI in understanding English. The findings from this experiment indicates that the MI at school play a major role compared to the region the student hailed from, as the rural English medium students performed better compared to urban regional language students. A reasonable gap is noticed between rural regional language students and urban English medium students. The cut-off marks, economical and educational background of their families and the gender of the student have minimal impact.

Hence it can be understood that the hypothesis of urban / rural exposure to the student having greater impact than the primary MI at school is reversed and the primary MI has greater impact on the performance in an English test, than the urban / rural exposure.

III. IMPACT OF MI IN UNDERSTANDING TECHNICAL CONTENT

The hypothesis of this survey experiment is that the primary MI at school has greater impact than the urban / rural exposure to the student, while understanding the technical contents given to them in English.

A. Participants

The same set of 219 participants who were tested in the previous experiment participated in this experiment as well.

B. Evaluation Procedure

The participants were given technical content through e-Learning engine which shows uniform content to all the users. At the end of the learning session, they were asked to answer questions from the contents they had learnt. They were given a total of 50 minutes to go through the contents and answer 20 questions about the contents.

C. Results Analysis and discussion

Table 2 summarizes the task performance of the impact of MI in understanding technical content in English. The findings from this experiment indicates that even in the case of technical content given in English using e-Learning, the MI at school played a major role compared to the region the student hailed from, as the rural English medium students performed better compared to urban regional language students.

TABLE 2

Medium of Study	Residence	No. of Students	Technical Content Test Score - Avg %	Range %
Regional	Rural	90	54.00%	35% -65%
Regional	Urban	58	56.12%	35% -70%
English	Rural	13	65.38%	45% -80%
English	Urban	58	74.48%	50% -100%

Table 2: Technical Test Performance

A reasonably good amount of gap noticed between rural regional language students and urban English medium students is a major technical criticality observed. Whereas the cut-off marks, economical and educational background of their families and the gender of the student have minimal impact.

The hypothesis that the primary MI at school has greater impact than the urban / rural exposure to the student, while understanding the technical contents is validated.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

IV. STATISTICAL REGRESSION MODEL

From Table 1 and 2 it is observed that the performance gap between English Medium students and Regional Language students are high to the extent of 10% difference in both English Test and Technical Test. When Rural Regional Language students' performance is considered, they are on par with Urban Regional Language students. This observation leads to the apparent decision that the performance of Rural / Urban English Medium students is far better than rest of the categories irrespective of the English Test or Technical Test.

In order to understand the gap, significant evidence is required. Hence, a Statistical Regression Model is developed to test the above hypothesis. In this study, performances of students in English and Technical tests were considered dependent variable. The cut-off mark scored by the student impacts the performance and is considered to be an independent variable. The impact of the economical and educational background of their families and the gender of the student are found to be negligible. So the model is developed as performance of a student is studied is a function of cut of mark.

Performance of Student = f(Cut-off Marks).

The impact of MI of the student and their place of study (PoS) on the performance in the test is analyzed by introducing them as dummy variables. Hence the analytical equation becomes:

$Y = a + b(X) + c(D_1) + e(D_2)$
Y = Performance on English Test
X = Cut-off Marks
D_1 = Dummy Variable (PoS)
D_2 = Dummy Variable (MI)

When the inputs are used, we get the values for a, b, c, e and also the coefficient of determination R^2 and F value. The coefficient of determination (R^2) explains the extent of explained variability of Y caused by X. The F denotes the significant effect of the ratio of variation due to the independent variables and the variation due to chance causes.

A. Regression (English Test – Effect of Cut-off Marks, PoS and Medium of Instruction)

This model analyses the effect of Cut-off Marks and the combined effect of PoS and MI on the performance in the English Test.

The values obtained for a, b, c and e is incorporated in the equation to form the following equation.

1. *Equation:* $Y = 14.252 + (-0.028)X + 1.248 (D_1) + 3.528 (D_2)$

The 't' value obtained for PoS, MI and for the Cut-off Marks and its level of significance are given below.

D_1 (PoS) = (2.517)**
D_2 (MI) = (6.673)***
X (Cut-off Marks)= $(-0.732)^{NS}$

*** - 1% Significance Level; ** - 5% Significance Level; * - 10% Significance Level; NS - No Significance

The regression equation is significant, since F = 25.513***. The capability of Cut-off Marks, PoS and MI is explaining the performance to the extent of 56.3% since $R^2 = 0.563$.

From the analysis it can be concluded that the effect of MI is significantly higher than that of PoS, even though the secondary plays a vital role.

B. Regression (Technical Test – Effect of Cut-off Marks, PoS and MI)

This model analyses the effect of Cut-off Marks and the combined effect of PoS and PI on the performance in the technical test.

The values obtained for a, b, c and e is incorporated in the equation to form the following equation.

1. *Equation:* $Y = 14.694 + (-0.046)X + 0.605 (D_1) + 1.685 (D_2)$

The 't' value obtained for PoS, PI and for the cut-off marks and its level of significance are given below.

D_1 (PoS) = (1.817)*
D_2 (MI) = (4.745)***
X (Cut-off Marks) = $(-1.782)^{*}$

*** - 1% Significance Level; ** - 5% Significance Level; * - 10% Significance Level; NS - No Significance

The equation is significant since F = 13.783*** and the capability of the independent variables to explain the variation in dependent variable is 46.1% since $R^2 = 0.461$.

From the above equation it can be concluded that the MI is playing significant role in contributing to the performance than the PoS, and the PoS has minimal impact.

V. PROPOSED ADAPTIVE E-LEARNING MODEL

Addressing the limitations of current e-learning platforms, we propose a model for adaptive e-learning. The fundamental concept of this model is that authors

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

create a bunch of individual learning objects, which are combined to form personalized courses according to each learner's capability. According to this model, learners are characterized based on their primary MI at school and their knowledge level of English.

The proposed AeL is developed using Moodle Learning Management System. Voluntary input obtained from the learner and assessment results are used to personalize the content offered to the learner.

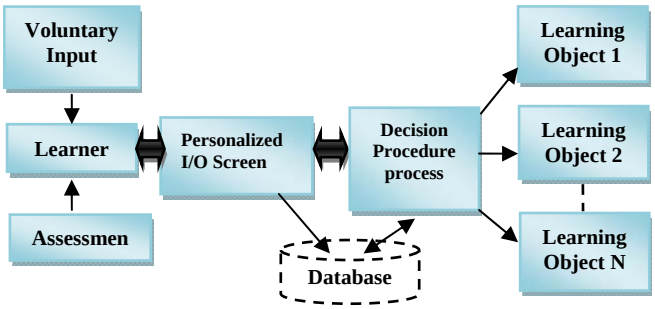


Fig: AeL Framework

There are various modules with different functionality in the proposed adaptive e-Learning model (Figure 1). The following information is voluntarily obtained from the learner: Age, Educational qualification, Urban / Rural, MI at School, familiarity with English language, gender, etc. Prior to start of a course the learner's proficiency of English is tested. On completion of the course, a post-test is given to assess the knowledge level gained.

Input is obtained for the learner through the "Personalized Input-Output Screen" and recorded in the "Database". The Database acts as the central repository on the Web for the Adaptive e-Learning. Based on the information available in the Database about the learner, the "Learning Objects" to be presented is automatically determined by the "Decision Procedure Process" and presented to the Personalized Input-Output Screen. The lessons of a course are split into various learning objects and they can be customized to the regional language students as follows:

- Prior to start of the actual content, present additional content for the regional language medium students to understand: (i) Phrases and Idioms, (ii) Grammar and (iii) Phonetics in English.
- The learning objects will display the regional language translation of English words on mouse over to a particular word.
- Present the learning objects comprising of glossary of English words and technical terms in the regional language. As and when the regional language students don't understand a certain word, they can refer to the glossary and understand its meaning in regional language.
- Present additional pages of course content for better understanding.

The Learning Objects can be presented in any permutation and combination as needed. For example, a course in Computer Networks with extensive content can be created. Depending on whether the learner's primary MI is English / Regional Language, and is an Undergraduate / Post-graduate / Research Scholar the lessons will be personalized.

The above indicated AeL platform has been developed using Moodle and implemented.

VI. ADAPTIVE E-LEARNING AS A TOOL TO OVERCOME THE MI BARRIER

The hypothesis of this experiment is that AeL can be successfully used to bridge the gap between regional language students and English medium students, while understanding the technical contents given to them in English. The same set of participants who were tested in the previous experiment participated in this experiment as well.

TABLE 3

Group	Course	No. of Students	Test Score – Avg %	Range %
Group-1 (English)	Learning Object-1 (English Only)	71	70.28%	45% -85%
Group-2 (Tamil)	Learning Object-1 (English Only)	74	40.81%	40% -75%
Group-3 (Tamil)	Learning Object-2 (Help in Regional Language)	74	60.74%	30% -55%

Table 3: Physics Test Performance – Adaptive e-Learning

A. Participants

The same set of heterogeneous group of 219 Undergraduate first year engineering students both from Urban and Rural backgrounds (116 Urban and 103 students hailing from Rural areas) and with a combination of regional language and English as their primary MI at school (148 Regional Language and 71 English medium students), but with similar pre-requisite subject knowledge.

B. Evaluation Procedure

The participants were divided into 3 groups. The Group-1 comprises of 71 students from English medium, the rest of the 148 students from regional language medium were equally divided into Group-2 and Group-3 irrespective of their Urban and Rural backgrounds. The Course was divided into 2 Learning Objects. The Learning Object-1 comprises of 15 pages of Physics subject content in English only. Learning Object-2 comprises of the same Physics subject content in English

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

with the regional language translation of English words on mouse over to a particular word and a glossary of English words and Technical terms in the regional language.

The Group-1 and Group-2 participants were shown Learning Object-1 and Group-3 participants were shown Learning Object-2. All participants were given a total of 60 minutes to go through the contents and answer 20 questions about the contents.

C. *Results and Discussion*

Table 3 summarizes the task performance of the impact of adaptive e-Learning in improving the performance. The findings from this experiment indicates that the performance gap between the English medium students and Regional language students can be narrowed down by translating the English content in Regional language as and when needed. Thus adaptive e-Learning enhances the performance of the students.

VII. STATISTICAL PAIRED T-TEST

From Table 3 it is observed that the performance gap between English Medium students and Regional Language students narrowed down by translating the English content in Regional language on mouse over to a particular word and by offering a glossary of English words and Technical terms in the regional language using adaptive e-Learning.

In order to understand the impact, significant evidence is required. Hence, a paired t-test is used to validate the impact of adaptive e-learning in improving the performance.

A. *Paired t-test (Regional medium students with English only content and Regional medium students with help in Regional language)*

This model analyses the gap in performance among the Regional language students when the content is given to them only in English and when help in Regional language is provided.

TABLE 4

	Regional Medium – Old Methodology (Contents in English only)	**Regional Medium – New Methodology (Help in Regional Language)**
Mean	8.162162162	12.14864865
Variance	13.89115143	7.635135135
Observations	74	74
df	73	
t Stat	-7.488277111	
P(T<=t) one-tail	6.41×10^{-11}	

The above table shows the analysis of the performance of Regional language students before and after the new methodology is introduced. From the analysis the mean score of Regional language students performance when contents are given in English with help in Regional language is 12.15 which is much higher than the average performance of 8.16 when the contents are given only in English. Further analysis done for finding statistical evidence for significant difference using paired t–test indicated the performance is significantly different because t = -7.49 @ 73 df (degrees of freedom) and is significant at 0% significance level.

B. *Paired t-test (English medium students and Regional medium students with help in Regional language)*

This model analyses the gap in performance among the English medium students with English only contents and Regional language students when the content is given to them in English with help in Regional language.

TABLE 5

	English Medium (Contents in English Only)	**Regional Medium – New Methodology (Help in Regional Language)**
Mean	14.15493	12.14865
Variance	14.36137	7.635135
Observations	71	74
df	143	
t Stat	3.65	
P(T<=t) two-tail	0.000363	

The above table shows the analysis of the performance of students with respect to the help related to appropriate MI. The mean score of students from English medium is equal to 14.15 and the mean score of Regional language students performance when contents are given in English with help in Regional language is 12.15 and hence the difference is narrowed down to the extent of 4 points approximately in the mean value when help is provided in regional language. The further analysis done for finding statistical evidence for significant difference using t–test indicated the performance is significantly different because t = 3.65 @ 143 df (degrees of freedom) is significant at 0% significance level. Even though when compared to the earlier level it has come down to a greater extent, the analysis indicate still there is a scope for improvement which can be addressed through other means.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

CONCLUSIONS AND FURTHER WORK

The findings from the first two experiments indicated that when content is offered in English, the MI at school is more significant compared to the region the student hailed from. The primary objective of this research is to bridge the gap between the two extreme sets of sample so that regional language students will perform as good as English medium students. This has been achieved by developing an adaptive e-Learning model (as detailed above) to personalize the content based on the profile of the student and other relevant factors. The finding from the third experiment indicated that the performance gap between the English medium students and Regional language students narrowed down significantly by translating the English content in Regional language on mouse over to a particular word and by offering a glossary of English words and Technical terms in the regional language.

Future research can be expanded to include for example, adaptive e-Learning content for students from non-English speaking countries to perform on par with students from English speaking countries, when the contents are offered to them in English.

REFERENCES

[1] Toshie Ninomiya, Ken Nakayama, Miyuki Shimizu, Fumihiko Anma and Toshio Okamoto: The Design of e-Learning Environment Oriented for Personalized Adaptability, Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007).

[2] Feng-Hsu Wang: A Contextualization Method of Browsing Events in Web-based Learning Content Management System for Personalized Learning, Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)

[3] Zhou Xiang: Design and implementing An integrated Learning Content Management System, 0-7695-3090-7/08 © 2008 IEEE

[4] Norman W.Y. Shao, Stephen J.H. Yang, Addison Y.S. Sue: A Content Management System for Adaptive Learning Environment, 0-7695-2031-6/03 © 2003 IEEE

[5] James L.Stapleton, H.Joseph Wen, Dave Starrett and Michelle Kilburn: Generational differences in using online learning systems, Human Systems Management 26 (2007) 99-109, IOS Press.

[6] Amal Al-Dujaily*, Hokyoung Ryu: A study on personality in designing adaptive e-learning systems, 978-0-7695-3167-0/08 © 2008 IEEE.

[7] Amal Al-Dujaily*, Hokyoung Ryu: A Relationship between e-Learning Performance and Personality, Proceedings of the Sixth International Conference on Advanced Learning Technologies (ICALT'06).

[8] Daniela Reddig, Joyce Karreman, Thea van der Geest: Watch Out for the Preview: The Effects of a Preview on the Usability of a Content Management System and on the Users' Confidence Level, 978-1-4244-2086-5/08 ©2008 IEEE.

[9] Jurriaan Souer, Paul Honders, Johan Versendaal, Sjaak Brinkkemper: Defining Operations and Maintenance in Web Engineering: a Framework for CMS-based Web Applications, 1-4244-1476-8/07 ©2007 IEEE.

[10] Farah Habib Chanchary, Indrani Haque, Md. Saifuddin Khalid: Web-Usage Mining to Evaluate the Transfer of Learning in a Web-based Learning Environment, 0-7695-3090-7/08 © 2008 IEEE.

[11] Tsandilas, T., & Schraefel, m. c. (2005, April 2-7). *An Empirical Assessment of Adaptation Techniques*, Paper presented at the CHI, Portland, Oregon, USA.

[12] Learning about e-learning © 2006 Kookaburra Studios.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Image Retrieval using Row Means of Column Transformed Even and Odd Parts of Image with Walsh, Haar and Kekre Transforms

Dr. H.B.Kekre[1], Sudeep D. Thepade[2], Akshay Maloo[3]
[1]Senior Professor, [2]Ph.D.Research Scholar & Associate Professor, [3]B.Tech (CE) Student
Computer Engineering Department, MPSTME,
SVKM's NMIMS (Deemed-to-be University), Mumbai, India
[1]hbkekre@yahoo.com, [2]sudeepthepade@gmail.com,[3]akshaymaloo@gmail.com

***Abstract*— The paper presents an innovative content based image retrieval (CBIR) technique based on row mean of column transformed even and odd part of image. Here the performance of proposed CBIR technique is tested for different transforms like Walsh transform, Haar transform and Kekre transform. Instead of using all pixels of image as feature vector for image retrieval, row mean of transform applied on columns of odd and even part of image is used, resulting into better performance with much lower computations. The proposed CBIR techniques are tested on generic image database having 1000 images spread across 11. For each transform 55 queries (5 per category) were fired. Then the precision and recall for all queries are computed. While considering the relevance of result images for calculation of precision and recall the results of odd image part and even image part both are ORED. The average precision and average recall values of all queries gives the performance comparison of proposed CBIR methods as compared to considering full image data as feature vector or considering the simple row mean as feature vector. The results have shown the performance improvement (higher precision and recall values) with proposed methods compared to all pixel data of image at reduced computations resulting in faster retrieval. The comparison of transforms for CBIR performance gives Walsh transform surpassing the other two considered here.**

Keywords-content based image retrival (CBIR), row mean, Walsh transform, Haar transform, Kekre transform.

I. INTRODUCTION

The large numbers of images are being generated from a variety of sources (digital camera, video, scanner, the internet etc.) which have posed technical challenges to computer systems to store/transmit and index/manage image data effectively to make such large collections easily accessible. Image compression deals with the challenge of storage and transmission, where significant advancements have been made [1,4]. The challenge to image indexing is studied in the context of image database [2,6,7,10,11], which has become one of the most promising and important research area for researchers from a wide range of disciplines like computer vision, image processing and database areas. The thirst of better and faster image retrieval techniques is increasing day by day. Some of important applications for CBIR technology could be identified as art galleries [12,14], museums, archaeology [3], architecture design [8,13], geographic information systems [6], trademark databases [21,22], weather forecast [5,21], medical imaging [5,18], criminal investigations [23], image search on the Internet [9,19,20].

A. Content Based Image Retrieval

In literature the term content based image retrieval (CBIR) has been used for the first time by Kato et.al. [4], to describe his experiments into automatic retrieval of images from a database by colour and shape feature. The typical CBIR system performs two major tasks [16,17]. The first being feature extraction (FE), where a set of features, called feature vector, is generated to accurately represent the content of each image in the database. The second task is similarity measurement (SM), where distance between the query image and each image in the database using their feature vectors is used to retrieve the top "closest" images [16,17]. For feature extraction in CBIR there are mainly two approaches [5] feature extraction in spatial domain and feature extraction in transform domain. The feature extraction in spatial domain includes CBIR techniques based on histograms [5], BTC [1,2,16], VQ [21]. The transform domain methods are widely used in image compression, as they give high energy compaction in transformed image [17,24]. So it is obvious to use images in transformed domain for feature extraction in CBIR [23]. But taking transform of image is time consuming, this complexity is reduced to a great extent by the proposed technique. Reducing the size of feature vector using pure image pixel data in spatial domain only and till getting the improvement in performance of image retrieval is the theme of the work presented. Many current CBIR systems use Euclidean distance [1-3,8-14] on the extracted feature set as a similarity measure. The Direct Euclidian Distance between image P and query image Q can be given as equation 1, where Vpi and Vqi are the feature vectors of image P and Query image Q respectively with size 'n'.

$$ED = \sqrt{\sum_{i=1}^{n} (Vpi - Vqi)^2} \qquad (1)$$

II. IMAGE TRANSFORMS

The various transforms [5,24] used for proposed CBIR techniques are discussed below:

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A. Walsh Transform

The Walsh transform matrix [5,18] is defined as a set of N rows, denoted Wj, for j = 0, 1, ... , N - 1, which have the following properties:

- Wj takes on the values +1 and -1.
- Wj[0] = 1 for all j.
- Wj xWkT=0, for j k and Wj xWkT =N, for j=k.
- Wj has exactly j zero crossings, for j = 0, 1, ..., N-1.
- Each row Wj is either even or odd w.r.t. to its midpoint.

B. Haar Transform

This sequence was proposed in 1909 by AlfrédHaar. Haar used these functions to give an example of a countable orthonormal system for the space of square-integrable functions on the real line [24]. The study of wavelets, and even the term "wavelet", did not come until much later. The Haar wavelet is also the simplest possible wavelet. The technical disadvantage of the Haar wavelet is that it is not continuous, and therefore not differentiable. This property can, however, be an advantage for the analysis of signals with sudden transitions, such as monitoring of tool failure in machines.

C. Kekre's Transform

Kekre's transform matrix is the generic version of Kekre's LUV color space matrix [1,18]. Kekre's transform matrix can be of any size NxN, which need not have to be in powers of 2 (as is the case with most of other transforms). All upper diagonal and diagonal values of Kekre's transform matrix are one, while the lower diagonal part except the values just below diagonal is zero.

Generalized NxNKekre's transform matrix can be given as:

$$K_{NxN} = \begin{bmatrix} 1 & 1 & 1 & .. & 1 & 1 \\ -N+1 & 1 & 1 & .. & 1 & 1 \\ 0 & -N+2 & 1 & .. & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & .. & 1 & 1 \\ 0 & 0 & 0 & .. & -N+(N-1) & 1 \end{bmatrix} \quad (2)$$

The formula for generating the term K_{xy} of Kekre's transform matrix is:

$$K(x,y) = \begin{cases} 1 & ,x \le y \\ -N+(x-1) & ,x = y+1 \\ 0 & ,x > y+1 \end{cases} \quad (3)$$

III. Proposed Technique

The CBIR technique given in [24] has been extended here. In [24] the feature vector considered for CBIR is row mean of column transformed image. Here first the odd part and even part of image are obtained using it's mirror image. Then transform is applied on each column of obtained images to calculate feature vector, image retrieval is done using both feature vectors (F1 & F2). The obtained results are combined using OR operator to obtain the final results. Figure 1 shows the feature extraction in proposed CBIR technique with row mean [25] of transformed image columns.

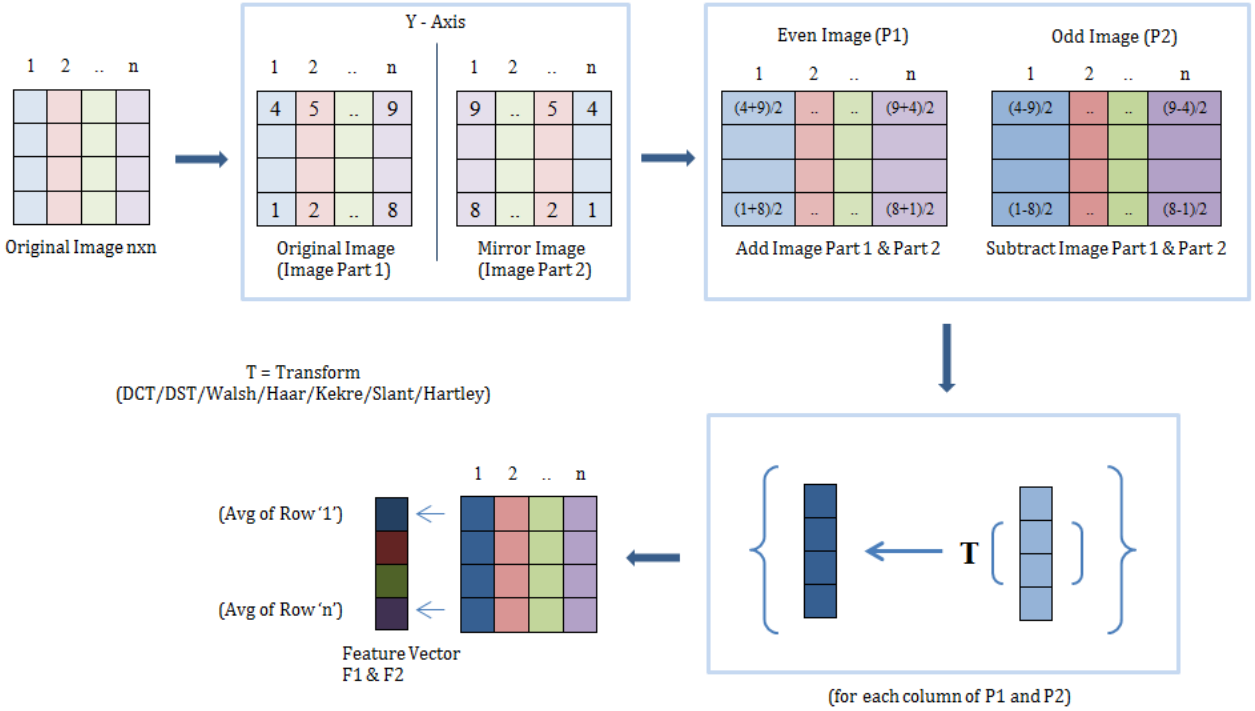


Figure 1 Feature Extraction in Proposed CBIR Technique using row means of column transformed even and odd part of image

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

R	R	N	R	N	R	Sorted Image Index for Even Part (P1)
R	R	N	N	R	R	Sorted Image Index for Odd Part (P2)
R	R	N	R	R	R	Result Image Index after using OR operator

Figure 2 How OR operator is used in proposed CBIR technique

The feature extraction steps for proposed image retrieval method can be given as follows.

1. Calculate the mirror image across Y-axis of the query image.
2. Using the image and its mirror image to calculate the Even Image (P1) and Odd image (P2). When you add both P1 and P2 you get the original image back.
3. Apply transform T on the column of P1 and P2 images of size NxN (I_{NxN}) to get column transformed images of the same size (cI_{NxN})

$$cI_{NxN}\text{ (column transformed)} = [T_{NxN}]\,[I_{NxN}] \quad (4)$$

4. Calculate row mean of column transformed image to get feature vector of size N (instead of N^2). We get two feature vectors for Even and Odd part (F1 & F2).
5. Then Euclidean Distance is applied and the results are sorted in ascending order for both feature vectors. The obtained results are combined using OR operator as seen in figure 2, to calculate precision and recall for the proposed technique.

The proposed technique is applied for three different transforms to see where the best performance is obtained. The results are compared with applying transform on full image and row mean on image (without column transform).

The figure 2 shows how OR operator is used in proposed CBIR techniques. Here the 'R' is relevant image retrieved and 'N' is non-relevant image retrieved. These relevancies are then considered to compute precision and recall values for all queries.

IV. Implementation

A. The Platform

The implementation of the proposed CBIR techniques is done in MATLAB 7.0 using a computer with Intel Core 2 Duo Processor T8100 (2.1GHz) and 2 GB RAM.

B. Database

The CBIR techniques are tested on the image database [15] of 1000 variable size images spread across 11 categories of human being, animals, natural scenery and manmade things. Figure 3 shows sample image of generic database.

Figure 3 Sample Images from Generic Image Database
[Image database contains total 1000 images with 11 categories]

C. Precision/Recall

To assess the retrieval effectiveness, we have used the precision and recall as statistical comparison parameters [1,2] for the proposed CBIR techniques. The standard definitions for these two measures are given by following equations.

$$\text{Precision} = \frac{Number\_of\_relevant\_images\_retrieved}{Total\_number\_of\_images\_retrieved} \quad (4)$$

$$\text{Recall} = \frac{Number\_of\_relevant\_images\_retrieved}{Total\_number\_of\_relevent\_images\_in\_database} \quad (5)$$

V. Results and Discussions

For testing the performance of each proposed CBIR technique, per technique 55 queries (5 from each category) are fired on the database of 1000 variable size generic images spread across 11 categories. The query and database image matching is done using Euclidian distance. The average precision and average recall are computed and are plotted against number of retrieved images. The crossover point of precision and recall gives important performance measure for image retrieval techniques. Higher the crossover point is better will be the performance of CBIR method.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

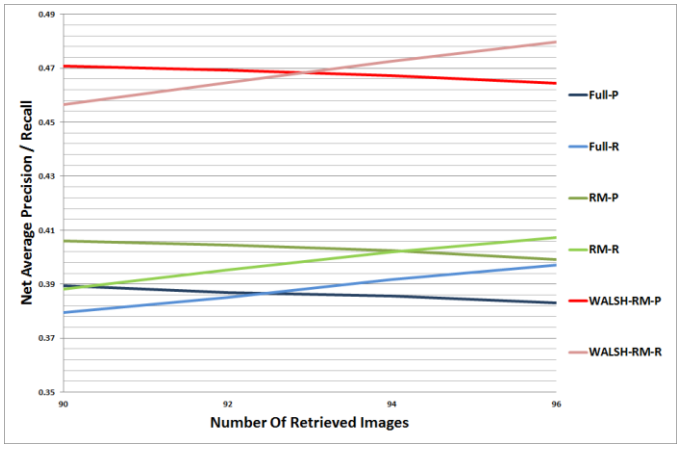


Figure 4 Crossover Point of Precision and Recall v/s Number of Retrieved Images using Walsh Transform

In Figure 4 the crossover points of CBIR using Walsh transform applied to the full image as feature vector (referred as 'Full'), simple row mean as feature vector (referred as 'RM') and the proposed CBIR technique with Walsh transform (referred as 'TRANSFORM-RM' here 'WALSH-RM') are shown. It can be noted from figure that proposed CBIR technique with Walsh transform gives best performance than the other discussed methods.

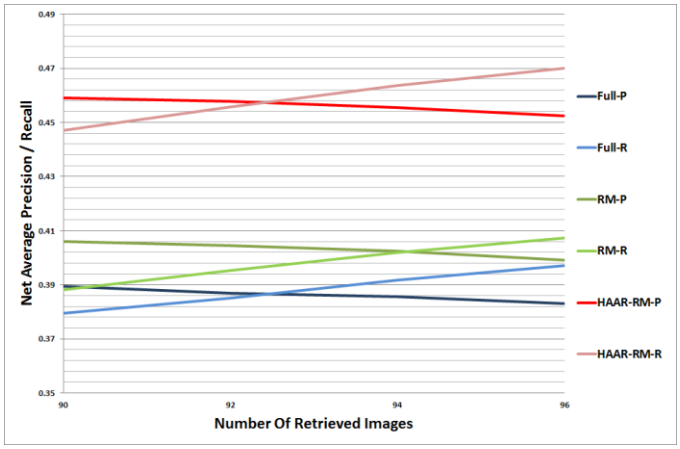


Figure 5 Crossover Point of Precision and Recall v/s Number of Retrieved Images using Haar Transform

The comparison of crossover points of CBIR with Haar transform of full image as feature vector (Full), row mean of image as feature vector (RM) and the proposed CBIR technique with Haar transform (HAAR-RM) are compared in figure 5. Also here the proposed CBIR method with Haar transform outperforms other image retrieval techniques.

Figure 6 gives the crossover points of CBIR using Kekre transform applied to the full image as feature vector (referred as 'Full'), simple row mean as feature vector (referred as 'RM') and the proposed CBIR technique with Kekre transform ('KEKRE-RM') are shown. It can be noted from figure that proposed CBIR technique with Kekre transform gives best performance than the other discussed methods.

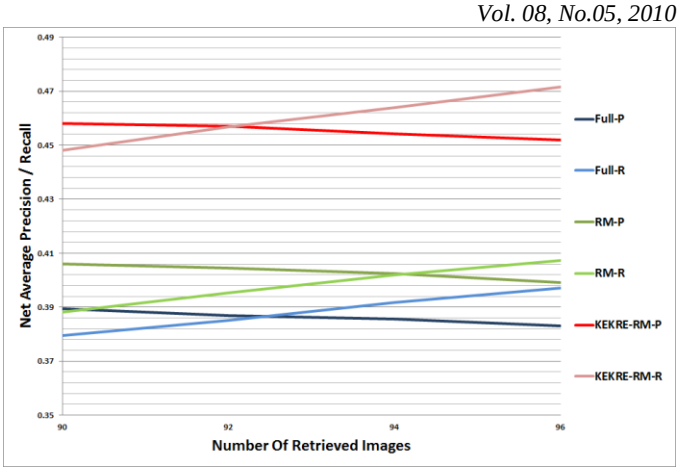


Figure 6 Crossover Point of Precision and Recall v/s Number of Retrieved Images using Kekre Transform

To compare the performance of the considered image transforms with each other for proposed CBIR techniques, the crossover points of proposed CBIR methods with individual transforms are shown together in figure 6. Here the best performance (highest precision-recall crossover point value) is given by Walsh transform. The performance ranking of transforms could be given as Walsh, Haar and then Kekre transform. From figure 7 also it can noted that proposed CBIR technique with all transforms is performing better than CBIR using full transformed image as feature vector or CBIR using row mean of image as feature vector.

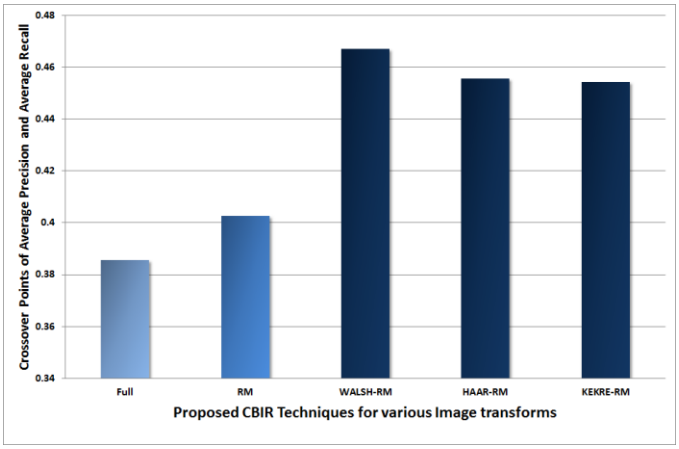


Figure 7 Crossover Point of average Precision and Recall v/s proposed technique for all transforms with Full Image

VI. CONCLUSION

The new CBIR method using row means of column transformed odd and even parts of the image with OR operator is proposed. The technique is tested using 55 queries fired on generic image database of 1000 images spread across 11 categories. The average precision and average recall values of these queries have proved the proposed CBIR technique better than taking complete transformed image as feature vector and also bettr than simple row mean of image as feature vector. The

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

proposed techniques are implemented using three mage transforms like Walsh transform, Haar transform and Kekre transform. The performance ranking of transforms for proposed CBIR method can be given as Walsh followed by Haar and then Kekre.

REFERENCES

[1] H.B.Kekre, Sudeep D. Thepade, "Boosting Block Truncation Coding using Kekre's LUV Color Space for Image Retrieval", WASET International Journal of Electrical, Computer and System Engineering (IJECSE), Volume 2, Number 3, pp. 172-180, Summer 2008. Available online at http://www.waset.org/ijecse/v2/v2-3-23.pdf

[2] H.B.Kekre, Sudeep D. Thepade, "Image Retrieval using Augmented Block Truncation Coding Techniques", ACM International Conference on Advances in Computing, Communication and Control (ICAC3-2009), pp. 384-390, 23-24 Jan 2009, Fr. Conceicao Rodrigous College of Engg., Mumbai. Is uploaded on online ACM portal.

[3] H.B.Kekre, Sudeep D. Thepade, "Scaling Invariant Fusion of Image Pieces in Panorama Making and Novel Image Blending Technique", International Journal on Imaging (IJI), www.ceser.res.in/iji.html, Volume 1, No. A08, pp. 31-46, Autumn 2008.

[4] Hirata K. and Kato T. "Query by visual example – content-based image retrieval", In Proc. of Third International Conference on Extending Database Technology, EDBT'92, 1992, pp 56-71

[5] H.B.Kekre, Sudeep D. Thepade, Akshay Maloo, "Performance Comparison of Image Retrieval Using Fractional Coefficients of Transformed Image Using DCT, Walsh, Haar and Kekre's Transform", CSC International Journal of Image Processing (IJIP), Vol. 4 , Issue 2, Computer Science Journals, CSC Press, Available online at http://www.cscjournals.org

[6] Minh N. Do, Martin Vetterli, "Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback-Leibler Distance", IEEE Transactions On Image Processing, Volume 11, Number 2, pp.146-158, February 2002.

[7] B.G.Prasad, K.K. Biswas, and S. K. Gupta, "Region –based image retrieval using integrated color, shape, and location index", International Journal on Computer Vision and Image Understanding Special Issue: Colour for Image Indexing and Retrieval, Volume 94, Issues 1-3, April-June 2004, pp.193-233.

[8] H.B.Kekre, Sudeep D. Thepade, "Creating the Color Panoramic View using Medley of Grayscale and Color Partial Images ", WASET International Journal of Electrical, Computer and System Engineering (IJECSE), Volume 2, No. 3, Summer 2008. Available online at www.waset.org/ijecse/v2/v2-3-26.pdf.

[9] Stian Edvardsen, "Classification of Images using color, CBIR Distance Measures and Genetic Programming", Ph.D. Thesis, Master of science in Informatics, Norwegian university of science and Technology, Department of computer and Information science, June 2006.

[10] H.B.Kekre, Tanuja Sarode, Sudeep D. Thepade, "DCT Applied to Row Mean and Column Vectors in Fingerprint Identification", In Proceedings of International Conference on Computer Networks and Security (ICCNS), 27-28 Sept. 2008, VIT, Pune.

[11] H.B.Kekre, Sudeep D. Thepade, Akshay Maloo "Performance Comparison for Face Recognition using PCA, DCT &WalshTransform of Row Mean and Column Mean", Computer Science Journals, International Journal oj Image Processing (IJIP), Volume 4, Issue II, May.2010, pp.142-155, available online at http://www.cscjournals.org/csc/manuscript/Journals/IJIP/volume4/Issue2/IJIP-165.pdf..

[12] H.B.kekre, Sudeep D. Thepade, "Improving 'Color to Gray and Back' using Kekre's LUV Color Space", IEEE International Advanced Computing Conference 2009 (IACC'09), Thapar University, Patiala, INDIA, 6-7 March 2009. Is uploaded and available online at IEEE Xplore.

[13] H.B.Kekre, Sudeep D. Thepade, "Image Blending in Vista Creation using Kekre's LUV Color Space", SPIT-IEEE Colloquium and International Conference, Sardar Patel Institute of Technology, Andheri, Mumbai, 04-05 Feb 2008.

[14] H.B.Kekre, Sudeep D. Thepade, "Color Traits Transfer to Grayscale Images", In Proc.of IEEE First International Conference on Emerging Trends in Engg. & Technology, (ICETET-08), G.H.Raisoni COE, Nagpur, INDIA. Uploaded on online IEEE Xplore.

[15] http://wang.ist.psu.edu/docs/related/Image.orig (Last referred on 23 Sept 2008)

[16] H.B.Kekre, Sudeep D. Thepade, "Using YUV Color Space to Hoist the Performance of Block Truncation Coding for Image Retrieval", IEEE International Advanced Computing Conference 2009 (IACC'09), Thapar University, Patiala, INDIA, 6-7 March 2009.

[17] H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Anant Shah, Prathmesh Verlekar, Suraj Shirke,"Energy Compaction and Image Splitting for Image Retrieval using Kekre Transform over Row and Column Feature Vectors", International Journal of Computer Science and Network Security (IJCSNS),Volume:10, Number 1, January 2010, (ISSN: 1738-7906) Available at www.IJCSNS.org.

[18] H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Anant Shah, Prathmesh Verlekar, Suraj Shirke, "Walsh Transform over Row Mean and Column Mean using Image Fragmentation and Energy Compaction for Image Retrieval", International Journal on Computer Science and Engineering (IJCSE),Volume 2S, Issue1, January 2010, (ISSN: 0975–3397). Available online at www.enggjournals.com/ijcse.

[19] H.B.Kekre, Sudeep D. Thepade, "Image Retrieval using Color-Texture Features Extracted from Walshlet Pyramid", ICGST International Journal on Graphics, Vision and Image Processing (GVIP), Volume 10, Issue I, Feb.2010, pp.9-18, Available online www.icgst.com/gvip/Volume10/Issue1/P1150938876.html

[20] H.B.Kekre, Sudeep D. Thepade, "Color Based Image Retrieval using Amendment Block Truncation Coding with YCbCr Color Space", International Journal on Imaging (IJI), Volume 2, Number A09, Autumn 2009, pp. 2-14. Available online at www.ceser.res.in/iji.html (ISSN: 0974-0627).

[21] H.B.Kekre, Tanuja Sarode, Sudeep D. Thepade, "Color-Texture Feature based Image Retrieval using DCT applied on Kekre's Median Codebook", International Journal on Imaging (IJI), Volume 2, Number A09, Autumn 2009,pp. 55-65. Available online at www.ceser.res.in/iji.html (ISSN: 0974-0627).

[22] H.B.Kekre, Sudeep D. Thepade, "Improving the Performance of Image Retrieval using Partial Coefficients of Transformed Image", International Journal of Information Retrieval**,** Serials Publications, Volume 2, Issue 1, 2009, pp. 72-79 (ISSN: 0974-6285)

[23] H.B.Kekre, Sudeep D. Thepade, Akshay Maloo "Query by Image Content Using Colour Averaging Techniques", Engineering journals, International Journal of Engineering, Science and Technology (IJEST), Volume 2, Issue 6, Jun.2010, pp.1612-1622, Available online http://www.ijest.info/docs/IJEST10-02-06-14.pdf.

[24] H.B.Kekre, Sudeep D. Thepade, Akshay Maloo, "Performance Comparision of Image Retrieval using Row Mean of Transformed Column Image ", International Journal on Computer Science and Engineering (IJCSE), Volume 2, Issue 5,2010

AUTHORS PROFILE

Dr. H. B. Kekre has received B.E. (Hons.) in Telecomm. Engineering. from Jabalpur University in 1958, M.Tech (Industrial Electronics) from IIT Bombay in 1960, M.S.Engg. (Electrical Engg.) from University of Ottawa in 1965 and Ph.D. (System Identification) from IIT Bombay in 1970 He has worked as Faculty of Electrical Engg. and then HOD Computer Science and Engg. at IIT Bombay. For 13 years he was working as a professor and head in the Department of Computer Engg. at Thadomal Shahani Engineering. College, Mumbai. Now he is Senior Professor at MPSTME, SVKM's NMIMS University. He has guided 17 Ph.Ds, more than 100 M.E./M.Tech and several B.E./B.Tech projects. His areas of interest are Digital Signal processing, Image Processing and Computer Networking. He has more than 320 papers in National / International Conferences and Journals to his credit. He was Senior

Member of IEEE. Presently He is Fellow of IETE and Life Member of ISTE Recently ten students working under his guidance have received best paper awards and two have been conferred Ph.D. degree of SVKM's NMIMS University. Currently 10 research scholars are pursuing Ph.D. program under his guidance.

Sudeep D. Thepade has Received B.E.(Computer) degree from North Maharashtra University with Distinction in 2003. M.E. in Computer Engineering from University of Mumbai in 2008 with Distinction, currently pursuing Ph.D. from SVKM's NMIMS, Mumbai. He has about than 07 years of experience in teaching and industry. He was Lecturer in Dept. of Information Technology at Thadomal Shahani Engineering College, Bandra(w), Mumbai for nearly 04 years. Currently working as Associate Professor in Computer Engineering at Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS University, Vile Parle(w), Mumbai, INDIA. He is member of International Association of Engineers (IAENG) and International Association of Computer Science and Information Technology (IACSIT), Singapore. He has been on International Advisory Board of many International Conferences. He is Reviewer for many reputed International Journals. His areas of interest are Image Processing and Computer Networks. He has about 79 papers in National/International Conferences/Journals to his credit with a Best Paper Award at International Conference SSPCCIN-2008, Second Best Paper Award at ThinkQuest-2009 National Level paper presentation competition for faculty and Best Paper Award at Springer International Conference ICCCT-2010.

Akshay Maloo is currently pursuing B.Tech. (CS) from MPSTME, NMIMS University, Mumbai. His areas of interest are Artificial intelligence, Image Processing, Computer Networks and Security. He has 13 research papers in International Conferences/Journals to his credit.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

IEEM PROGRAMMING PROCEDURE FOR DETECTING BOUNDARY OF CAROTID ARTERY

V.Savithri
Professor,Department of Computer Science
Mother Teresa Women's University
Kodaikanal, INDIA
Savi3_8@yahoo.co.in

Dr.S.Purushothaman
Principal, Sun College of Engineering and Technology
Kanyakumari, INDIA
dr.purushothaman.s@gmail.com

Abstract--This paper presents an IEEM programming procedure for use on noisy B-mode ultrasound images of the carotid artery. This programming procedure is based on Image Enhancement, Edge detection and Morphological operations in Boundary detection. This procedure may simplify the job of the practitioner for analyzing accuracy and variability of segmentation results. Possible plaque regions are also highlighted. A thorough evaluation of the method in the clinical environment shows that inter observer variability is evidently decreased and so is the overall analysis time. The results demonstrate that it has the potential to perform qualitatively better than applying existing methods in intima and adventitial layer detection on B-mode images.

Keywords- Artery, boundary detection , imaging, Ultrasonic, parallel programming.

I. Introduction

Over the last few years, image processing has been playing an increasingly important role in many scientific areas. This is due, among other reasons, to the ever-improving performance of computers that are now capable of quickly processing the characteristically large amounts of data produced by images. This processing is mostly oriented toward extracting either qualitative or quantitative information from object images. In particular, [1][2]precise dimensional characterization of objects through contact-less measurement techniques is a very important task in several environments such as industrial quality and process control and medical diagnosis.[4] A typical application field of medical image processing is in the diagnosis of atherosclerosis.[5]The atherosclerosis process is strongly linked to carotid thickening and plaques, whose presence can be clearly detected in artery longitudinal section images provided by ultrasound techniques.[6]The analysisof the carotid ultrasound represents, in fact, the most powerful instrument today available for predictingcoronary disorders in people over fifty .Images are taken by ultrasound equipment working at frequencies ranging from 1 to 20 MHz, and are obtained via a probe that has to be positioned on the patient's neck. [3]There are two phases to the measurement process: i) carotid image capturingand ii) thickness measurement.

Ultrasound plays an important role in diagnosis and illness and injury. The noninvasive imaging of different parts of the body has other applications in medical diagnosis, such as in tissue characterization and measurement of tissue motion.The measurement of tissue motion[9] is a broad category. It can include the analysis of the motion of physically active organs such as the heart and discrete structures such as cardiac valves and arterial walls. It can refer to the analysis of motion induced in passive tissues, such as liver or lung, due to active organs such as the heart or by external sources such as low frequency vibration or compression. The response of the tissue is a function its elasticity, which is directly related to the healthiness of the tissue. [12]The measurement of blood flow velocity is an important parameter in diagnosing vascular diseases such as venous thrombosis. Regardless of the particular medical application of ultrasound, all applications require the transmission and capture of radio-frequency (RF) ultrasonic signals. The signals must be processed in some way to extract the desired information. In the case of assessing tissue motion, Doppler ultrasound has been very popular, particularly in the measurement of blood flow. Practically all commercial ultrasound blood flow measurement systems utilize the frequency-domain Doppler technique. Doppler techniques have been around for a long time, and have been extensively covered in the literature. In addition to Doppler, however, time-domain methods of measuring tissue motion also exits, which have not been comprehensively reviewed.[17]Time-domain methods have potential advantages over Doppler techniques in many applications, and the use of time-domain based methods is becoming more and more widespread.

Due to[6] the huge amount of information intravascular**(**IVUS) images are increasing their role in the diagnosis and treatment of several diseases. Manual

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

segmentation is slow and lacks of objectivity.[10] Consequently, automatic segmentation and tracking of the vessel inner wall in IVUS images has been approached in several recent works.[8] The poor quality of the images suggests the use of techniques such as probabilities or fuzzy logic guiding an active contour to adjust the inner wall.

II. Materials and Methods

Problem Definition

The noise created during ultrasound scanning leads to difficulty in defining the boundary of the vessel. The image is further deteriorated by the occurrence of lipid rich plaque a poorly angled transducer during image acquisition. Difficult in highlighting plaque region.

BOUNDARY DETECTION

Ultrasonic Artery Images

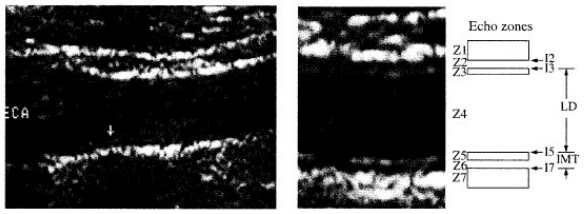


Figure 1(a) Carotid artery image. (b) Definition of echo zones and interfaces.

Figure. 1(a) shows a representative image of a carotid artery. The femoral artery has a similar appearance. The echoes in the region of interest can be schematically grouped into seven echo zones Z1–Z7 [Figure. 1(b)]. Previous studies [12], [13] have shown that the leading edge (upper side) of Z3, Z5, and Z7, denoted as I3, I5, and I7, can be mapped to the near-wall intima–lumen interface, the far-wall lumen–intima interface, and the far-wall media–adventitia interface, respectively. Consequently, the distancebetween I3 and I5 represents the LD and the distance between I5 and I7 is the far-wall IMT.With this understanding, the determination of ultrasonic measurement of the artery becomes equivalent to accurately detecting the echo boundaries I3, I5, and I7.

[14]The femoral artery has a similar appearance. The echoes in the region of interest can be schematically grouped into seven echo zones Z1–Z7 .[15][16] Previous studies have shown that the leading edge (upper side) of Z3, Z5, and Z7, denoted as I3, I5, and I7, can be mapped to the near-wall intima–lumen interface, the far-wall lumen–intima interface, and the far-wall media–adventitia interface, respectively. Consequently, the distance between I3 and I5 represents the LD and the distance between I5 and I7 is the far-wall IMT. With this understanding,[14 the determination of ultrasonic measurement of the artery becomes equivalent to accurately detecting the echo boundaries.

III. IEEM programming procedure

Image Enhancement:

Histogram equalization provides more visually pleasing results across a wider range of images. Image brightness[11] may be improved by modifying the histogram of the image.

Image Denoising:

An image is often corrupted by noise in its acquisition or transmission. Noise is any undesired information that contaminates an image. Noise appears in images from a variety of sources. The goal of denoising is to remove the noise while retaining as much as possible the important signal features. Denoising can be done through filtering. Filters reduces noise. Gaussian highpass filter helps to reduces a noise.

```
% H = guassian_filter(co,ro,fo);
H = 1 – H;
Out = Zeros(co,ro);
Outf = imf. * H;
Out = abs ( ifft2(outf));
Imshow(im), title ('Original image'), figure,
 Imshow((out)), title ('Filtered Image') figure,
 Imshow (H), title('2D view of B'), figure, surf(H),
 Title ('3D view of H')
```

Edge Detection:

The IEEM defines edges as Zero-crossings of second derivatives in the direction of the greatest first derivate. This works in multistage process (i) image is smoothed by Gaussian convolution (ii) 2D first derivate operator is applied to the smoothed image to highlight region of the image with high spatial derivatives. The effectiveness of this algorithm is determined by three parameters (i) width of the Gaussian kernel (ii) upper threshold (iii) lower threshold used by tracker.

Morphological operations for Boundary detection:

Morphological operations are very effective in the detection of boundaries in a binary image X. The following boundary detectors are widely used:

$Y = X – (X \theta B)$

$Y = (X \oplus B) = X$ or

$Y = (X \oplus B) - (X \theta B)$

where Y is the boundary image, operator θ denotes erosion operator $\oplus$ denotes dilation ' – ' denotes set theoretical subtraction.

```
%Boundary detector
Close all;
Clear all;
Clc;
a=imread('carotid.jpg');
b=[010;111;010];
a1=imdilate(a,b);
a2=imerode(a,b);
a3=a-a2;
a4=a1-a;
a5=a1-a2;
imshow(a)
figure,imshow(a1),title('Dilated Image')
figure, imshow(a2),title('Eroded Image')
```


http://sites.google.com/site/ijcsis/
ISSN 1947-5500

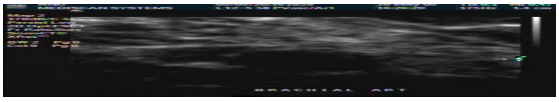
Segmentation of the intima-media region:

We introduce a new method for the segmentation of the imtima-media region in ultrasound images, which combines splines(for the adventitia detection), dynamic programming(dp), smooth intensity thresholding surfaces and a successful geometric active contour model and known for its accuracy, flexibility and robustness. Several image features are used in the segmentation. Human interaction is minimal. It is able to segment both near-end and far-end carotid walls; it supports to detect plaques of different sizes, shapes and classes.

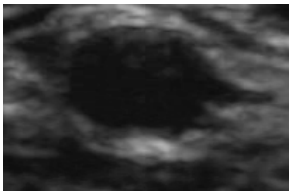
IV. Results and discussion

TABLE 1
READING VARIABILITY(%) WHEN MEASUREMENTS WERE PERFORMED BY THREE READERS BEFORE APPLYING IEEM PROCEDURE.

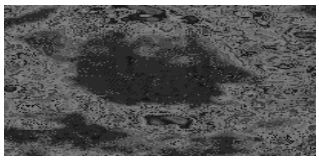
READER1 (accuracy)	READER2 (accuracy)	READER3 (accuracy)	AVERAGE (accuracy %)
82%	80%	84%	82%



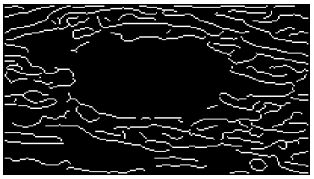
Original image



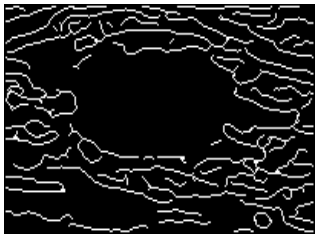
gray



indexed



G-canny



morphed

TABLE 2
READING VARIABILITY(%) WHEN MEASUREMENTS WERE PERFORMED BY THREE READERS AFTER APPLYING IEEM PROCEDURE.

READER1 (accuracy)	READER2 (accuracy)	READER3 (accuracy)	AVERAGE (accuracy %)
94%	94.5%	96%	94.5%

V. Conclusion

In conclusion, We have proposed a method based on IEEM programming procedure to automatically measure ultrasonic artery images. The human knowledge of the artery image is incorporated in the system, which makes the system capable in processing images of different quality. Human factors in the determination of the boundaries are reduced. Evaluation of the system shows reduced inter observer variability as well as overall analysis time. The automated artery boundary detection system can replace the old manual system in a clinical application environment.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Acknowledgement

We would like to thank members of Madras Medical College and New hope Scan center And Hospital , in the Department of Radiologist and Neurologist ,and Dr.P.Kamalakannan, offered valuable comments advices and for providing patients personal reports

References

[1] R. Chelappa *et al.*, "The past, present, and future of image and multidimensional signal processing," *IEEE Signal Processing Mag.*, vol. 15, pp. 21–58, Mar. 1998.

[2] D. H. Ballard and C. M. Brown, *Computer Vision.* Englewood Cliffs, NJ: Prentice-Hall, 1982.

[3] L. Angrisani, P. Daponte, C. Liguori, and A. Pietrosanto, "An image based measurement system for the characterization of automotive gaskets," *Measurement*, vol. 25, pp. 169–181, 1999.

[4] M. G. Bond and S. K. Wilmoth *et al.*, "Detection and Monitoring of Asymptomatic Atherosclerosisi in Clinical Trials," *Amer. J. Med.*, vol.86, (suppl 4A), pp. 33–36, 1989.

[5] N. M. El-Barghouty, T. Levine, S. Ladva, A. Flanagan, and A. Nicoladeis, "Histological verification of computerized carotid plaque characterization," *Eur. J. Vascular Endovascular Surg.*, vol. 11, pp. 414–416, 1996.

[6] F. De Man, I. De Scheerder, M.C. Herregods, J. Piessens and H. De Geest Role of Intravascular Ultrasound in Coronary Artery Disease: A new gold standart? Beyond Angiography. Intravascular Ultrasound State-Of-The-Art XX Congres of the ESC, Vol 1 (August 1998)

[7] D. Hausmann, Andre J.S. Lundkvist, Guy Friedrich, Krishnankutty Sudhir, Peter J. Fitzgerald and Paul G. Yock Lumen and Plaque Shape in Atherosclerotic Coronary Arteries Assesed by In EVO Intracoronary Ultrasound Beyond Angiography. Intravascular Ultrasound: State-Of-The-Art XX Congres of the ESC, Vol 1 (August 1998)

[8] F. Escolano, M. Cazorla, D. Gallardo and R. Rizo Deformable Templates for Plaque Thickness Estimation of Intravascular Ultrasound Sequences Pattern Recognition and Image Analysis. Preprints of the VI1 National Symp. On Patt. Recog. and Im. An.Vol 1 (April 1997)

[9] M.A.Bottalico, A.Starita, "EcoStudio:A computer tool to support carotid ultrasound images analysis,Engineering in Medicine and Biology Soc.,IEEE,pp.2428-2430,2000.

[10] Song Chun Zhu, Alan Yuille, Region Competition:Unzfiing Snakes, Region Growing, and BayesIMDL for Multiband Image Segmentation. IEEE Trans. Pattern An. Mach. Intelligence, Vol. 18, No 9 , (September 1996).

[11] Nobuyuki Otsu A Threshold Selection Method from Gray-Level Histograms. IEEE Trans. on Sys. Man and Cybernetics, Vol. SMC-9,Na 1, pp 62-65, (January 1979)

[12] I. Wendelhag, T. Gustavsson, M. Suurkula, G. Berglund, and J. Wikstrand, "Ultrasound measurement of wall thickness in the carotid artery. Fundamental principles and description of a computerized analyzing system," *Clin. Physiol.*, vol. 6, no. 11, pp. 565–577, 1991.

[13] P. Pignoli, E. Tremoli, A. Poli, and R. Paoletti, "Intimal plus medial thickness of the arterial wall: A direct measurement with ultrasound imaging," *Circulation*, vol. 74, pp. 1399–1406, 1986.

[14] R. E. Bellman and S. Dreyfus, Appled Dynamic Programming. Princeton, NJ: Princeton University Press, 1962.

[15] A. A. Amini, T. E.Weymouth, and R. C. Jain, "Using dynamic programming for solving variational problems in vision," IEEE Trans. Pattern Anal. Machine Intell., vol. 12, pp. 855–867, Sept. 1990.

[16] R. J. Kozick, "Detectinh interfaces on ultrasound images of the carotid artery by dynamic programming," *SPIE*, vol. 2666, pp. 233–241, 1996.

[17] W. Liang, R. Briwning, R. Lauer, and M. Sonka, "Automated analysis of brachial ultrasound time series," in *Proc. SPIE Conf. Physiol. Function Multidimensional Images*, vol. SPIE 3337, San Diego, CA, Feb. 1998, pp. 108–118.

Savithri Vedachalam , working as professor in Department of Computer Application in Hindu College, Pattabiram, Chennai,. She received her degrees, M.Phil.(C.S.) from Alagappa University and M.C.A. from Annamalai University, India.Her area of research includes Medical Imaging, Image Processing, Object tracking, 3-D Image analysis. She published more than 10 papers in national , International Conferences and Journals.

.

. Dr. S. Purushothaman is working as Principal and professor in Sun College of Engineering and Technology, Nagerkoil, India. He received his Ph.D from IIT Madras, M.E from Anna University Chennai and B.E from PSG College of Technology, Coimbatore His area of research includes Artificial Neural Networks, Image Processing and signal processing. He published more than 50 research papers in national and international journals.

A Scenario-Based Mobile Application for Robot-Assisted Smart Digital Homes

Ali Reza Manashty, Amir Rajabzadeh and Zahra Forootan Jahromi
Department of Computer Engineering
Razi University
Kermanshah, Iran
a.r.manashty@gmail.com, rajabzadeh@razi.ac.ir, zahra.forootan@gmail.com

***Abstract*—Smart homes are becoming more popular, as every day a new home appliance can be digitally controlled. Smart Digital Homes are using a server to make interaction with all the possible devices in one place, on a computer or webpage. In this paper we designed and implemented a mobile application using Windows Mobile platform that can connect to the controlling server of a Smart Home and grants the access to the Smart Home devices and robots everywhere possible. UML diagrams are presented to illustrate the application design process. Robots are also considered as devices that are able to interact to other object and devices. Scenarios are defined as a set of sequential actions to help manage different tasks all in one place. The mobile application can connect to the server using GPRS mobile internet and Short Message System (SMS). Interactive home map is also designed for easier status-checking and interacting with the devices using the mobile phones.**

Keywords- smart homes; mobile applications; remote home controls; automated digital homes; robot assisted at home; general packet radio service (GPRS); short message system (SMS); robot assisted at home; scenario based smart home.

I. INTRODUCTION

Smart homes are becoming more popular, every day a new home appliance can be controlled digitally. New wireless technologies also help the integration of remote controls into regular mobile devices so that users can control all the different appliances using a single device. As the number of devices in the home increase, it will become harder for manufacturers to adopt a universal standard for application controls.

Ease of access and use, is the main purpose of many remote controllers we now use for our devices. Their number is getting bigger and bigger each day, as a new device becomes remotely controllable. Speakers, air conditioners, lights, curtains, garage door, TVs and players are already being remote controlled. Now that every single part of our homes can be controlled remotely, we must think of a single remote control for all of the possible actions we want to take in the house.

Smart Home is not a new term for science society but is still far more away from people's vision and audition. This is because although recent various works has been done in designing the general overview of the possible remote access approaches for controlling devices [1], or in cases simulating the Smart Home itself [2], and designing the main server [3], the design and implementation of an off-the-shelf smart home remote control application has been limited to simply the computer applications and just in cases mobile [4] and web application development [5]. Nowadays people spent a noticeable amount of time in transportation, without having access to their PCs or having hard time accessing their laptops; instead, they are constantly using their cell-phones/PDAs. Because of this, we designed and implemented a mobile application that can be connected to a server where other access routes such as web application and local windows application are also available in there.

An important question regarding the problem is whether developing the web application can take the place of mobile phone applications, due to availability of web sites through GPRS and WiMax wireless internet? The answer to this question is that even though we can access our home control system through mobile wireless internet and use of current mobile browsers, which are now no less powerful than PC browsers, they cannot access GSM messaging systems such as SMS, MMS and so on. On the other hand, simultaneous accessing to mobile internet services (e.g., GPRS) for loading a complete webpage, is still expensive, so there will not be any need for designing the home control schedules and rules [5] online; instead, a temporary connection will do the information updating while using the mobile application offline.

In this paper, we present design of a mobile application for accessing and controlling the smart home control systems. We also show an implementation overview using Windows Mobile platforms and C# language as well as the general outline of the system. The term scenario is defined in this paper and the way of specifying robot-assisted tasks is also described.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

II. System Architecture

A. Preface

The Smart Home system usually consists of several devices scattered around the house that are linked together using a wired or wireless network. A computer system acts as the server in a node which controls all the information exchange through the network. The server system must have a device manager -middleware- [3] that assists the main application by connecting all the different device controllers through a single interface or the least interfaces possible.

B. Devices

The various types of devices in this house can be divided into three categories:

1. Actuator devices, e.g., alarms, lights and doors
2. Sensor devices, e.g., heat, movement and healthcare
3. Actuator/Sensor devices, e.g., robots, air conditioners

All the devices (whether sensor or actuator) can express their status using their controllers, such as whether they are on or off or the job they are currently involving (e.g. closing the door).

A device driver might also be needed for more complex devices; because each device might be anything from a lamp to a home robot.

C. Robots

Robots can be regarded as an actuator/sensor device, yet with wider range of abilities. Robots can be used in future homes more efficiently than the devices. This can be done by using robots for many actions that different devices can do independently. Robots (like Joy steward robot [6]) have many different capabilities that can be used in a smart home control system. We can use robots in many places where devices cannot be controlled independently, like moving and cleaning other device. Using the robots can eventually lower the number of controllable devices to as low as the number of robots in a home.

Another advantage of using robots instead of controllable devices is that there will not be any need for connecting different devices in predefined positions in the home; instead robots can move to any place in the home, controlling and monitoring all the possible devices a home can contain. This will help integrating the home with a smart robot far faster and easier than installing the controller for different devices. On many devices, installing controllers may also be difficult or even impossible.

Because current commercial robots, like cleaning bots, are able to do a limited set of actions a combination of controllable devices and robots is recommended. So for controlling devices using robots, we must make a three level scheduling schema in which, first the robot, then the device and then the corresponding device action must be selected.

D. Subsystems

The server application has the capability of adding newly installed devices and providing the appropriate controlling methods. The controlling signal and status controlling schema for all these devices is available through a general interface. We refer any of these devices, as objects. So the application must be object-independent while the device manager is closely in contact with all these objects through their device drivers and appropriate connections (e.g., cable or wireless Ethernet) as King et al. designed Atlas platform as a middleware in this field [7]. Using the object-independent programming interface, we can extend the controlling methods to any further possible ways such in the web application, mobile application and telephone line controller; easily without the need of changing the application codes. The subsystems in this smart home controlling environment is illustrated is Fig. 1.

The database of the system is playing an important role. The connection of the web application and mobile application to the home control system will be through the information in the database.

There are several parts of the system that must be designed and linked together. From the subsystems showed in Fig. 1, we are going to design and implement the mobile application, because, as mentioned in the Introduction section, the other systems had been designed before.

Before designing the application itself, we must design the server connection schema of the system, in which the connections of the different parts are modeled.

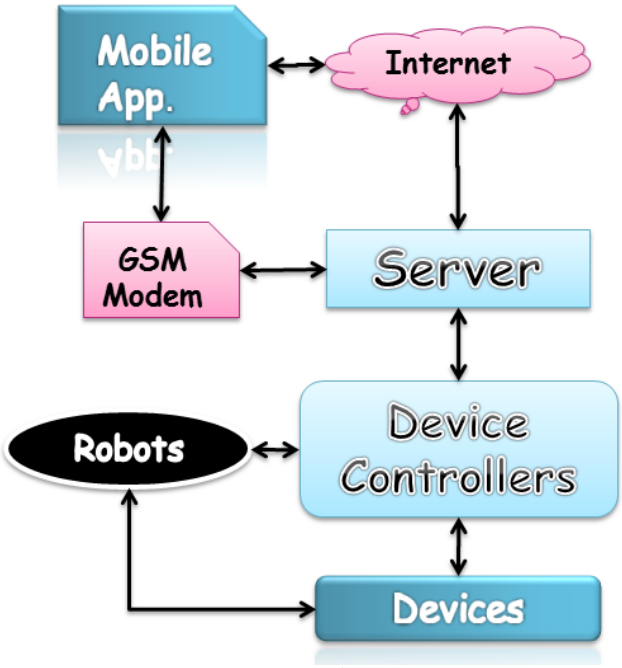


Figure 1. The subsystems overview of a robot-assisted Smart Home controlling system using a mobile application.

E. Server

The server is actually the computer system in the smart home that contains the windows application and device

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

manager. The server retrieves the sensor information on regular intervals and updates the database. This interval is different for vital and non-vital devices. For example, elderly and disabled person's health monitoring sensors [8, 9] data must be updated at least every second. The burglary detection system can be updated every few seconds but the temperature and light sensors might be checked within minutes. These categories help energy-saving schemas to be applicable.

This database also contains information about the devices, scenarios, rules, user access and other policies. Other remote control applications, such as web and mobile application, retrieve sensor and device information from this database and update the scenarios and rules accordingly so each time server updates the database sensor information, it also checks the changes applied to rules and scenarios and perform the actions necessary.

The web application can connect directly to the database, but because of functional restriction, mobile devices cannot connect directly to the server database and update the information like regular data connections. So here we face two important issues. The first issue is where to keep the database, so remote application can have continuous access to it? The answer to this question is completely related to the web application and security. Web applications can be either hosted in the Smart Home server or a host and domain reseller server. Due to security measures, we decided to make the web server and the database all in same place in the home server. This requires a static IP for the house, which is not a problem as security comes first. So the location of the web application is in the Smart Home server and the database is shared between the web and windows application. The second issue is the connection of the mobile application to database for retrieving and updating the information. As mentioned before, current mobile applications may not have the required memory and libraries required to establishing a direct connection to the server, so the only way we can exchange information is through http web servers using GPRS (small data packets can be sent through SMS to the GSM modem attached to server). One possible way to do this task is using web requests. Web requests are parameters send to web site using the "?" operator after the webpage name (e.g., www.test.test/login.aspx?user=admin&&pass=123456). After processing the web page requests, it can be identified that a request has been set using the mobile application (using the appropriate web ***requests***), and the page ***response*** will be changed according to the ***request***.

The web response page is regularly the html content of the website, but using ASP.Net application, before the page can be loaded, according to the web ***requests*** we can send limited lines of information instead of the whole html page. So we can use this feature to exchange information between the mobile application and a simple aspx page we already developed in our web server.

Now we described the overall server behavior. From here on, we focus on the mobile application design and implementation.

III. SCENARIO-BASED MOBILE APPLICATION DESIGN

A. Scenarios

Scenarios are also very impressive in this design. Scenarios are list of actions that consists of other different scheduled tasks or scenarios. A scheduled task is simply a scenario with only a single task.

These scenarios make it easy for people saving the list of actions for further use, in addition to design multiple actions to be done in a single scenario. Later the scenarios can be enabled/disabled in the scenarios list or be used in another scenario too. Cheng, Wang and Chen proposed a reasoning system for smart homes that is also scenario based [10].

Here with robots able to do different actions to other devices, it may be necessary also to select the robot in first place and the device and then the action the robot must do with the device. Since different robots can do different things on different devices, when the desired robot is selected, the appropriate devices and actions are available for the robot. For example a robot can move things around the house, the other one can clean different things, one may cook and another one will check the different rooms of the house for intruders and phenomena (e.g., catching fire and breaking of water pipes).

An example of a scenario of the scheduled tasks is as below:

Scenario name: Watering Plants

A. Sprinkler 1: on @ 5:00 AM

B. Sprinkler 2: on @ 5:30 AM

C. Sprinkler 1: off @ 7:00 AM

D. Sprinkler 2: off @ 9:00 AM

An important fact we must consider is that although many actions on devices are single state tasks, another tasks might need some parameters. For example when we want to set the temperature of an air conditioner system, we might also need to pass the desired temperature to the device. This is more important and complex when using special robots. For example when a robot is capable of moving or cleaning the objects, we must use a term for the action that both indicate the action and the subject of the action. We show parameters of an action in parentheses.

Only one time a user sets the actions in the scenario and then he or she can use it several times after defining it. Another advantage of using scenarios is that we can use other scenarios in the scenario we are currently defining. Example of scenarios involving robots and other scenarios is as below:

Scenario name: Clean Home

A. Cleaning robot: Clean (Bathtub) @ Now

B. [Gather Dishes] @ 10:00 AM

C. Home robot **à** Washing machine: on @ 10:05 AM

D. Cleaning robot: Clean (Saloon) @ 10:05 AM

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The above *Clean Home* scenario consists of three scheduled tasks and one scenario called *Gather Dishes*. Scheduled time of the scenarios will be overridden when a new time is set (like task *B* in above scenario). Scenarios used in another scenario are placed in brackets ([ ]). *Washing machine* and *Cleaning robot* are considered as devices here. The words followed by them after ":" are the actions they must take. When a robot is used to do an action of a device, the first item of the scenario action will be the name of the robot doing the action followed by an arrow (**à**) (e.g., Home robot will turn the Washing machine on, not the machine itself). In this scenario, *Clean* action of the *Cleaning robot* is parameterized so that the robot knows how to come along with the parameter.

Another example is the *Gather Dishes* scenario that was already used in Clean Home scenario and contains parameterized actions of the *Mover robot*:

Scenario name: Gather Dishes

A. Mover robot: GoTo (Saloon) @ Now

B. Mover robot: PickUp (Dishes) @ In 2 Minutes

C. Mover robot: GoTo (Kitchen) @ In 5 Minutes

D. Mover robot: PutInto (WashingMachine) @ In 6 Minutes

E. Mover robot: GoTo (DefaultPosition) @ In 7 Minutes

B. Use-Cases

The main server computer, which is located in the smart home area, is loaded with the windows application that gives the administrator user a comprehensive set of options and capabilities. The user can add and manage devices and robots in the application (of course if hardware procedures had been proceeded previously), design the home top view plane using graphical tools and icons, manage user access controls (e.g., define access limits for children), define policies of remote access (e.g., authenticated phone numbers), define rules (conditions to be checked and actions taken if the predefined criteria is met), define scenarios and check the current status of devices and robots (Fig. 2)

In the other hand, mobile application user can do the most common and important tasks, but not all the functionalities, as illustrated in Fig. 2. This limitation is mainly because of limits of a mobile application implementation and security factors. For example the mobile application user cannot design the home top view plane that graphically shows the current status of the home, but can simply view it from his/her cellphone.

Mobile applications can use advantages of cellphones such as built-in microphone and color display that usual remote controls suffer from having it. This will make mobile applications capable of live streaming of cameras within the house and implementing speech recognition for ease of access [11].

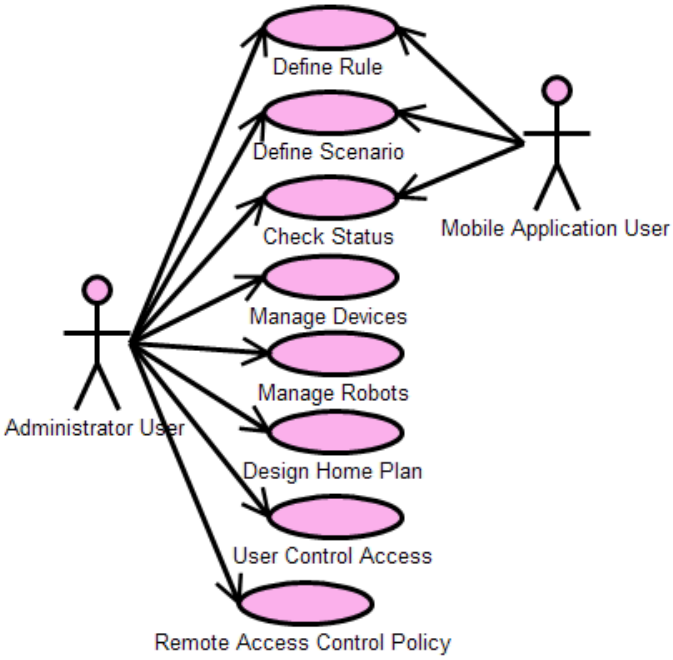


Figure 2. Use-cases of the Administrator and the Mobile user

C. Connections to Server

As mentioned earlier, we must connect to the database using a web server. The web application can have most of the capabilities and access levels of the windows application of the server but with some restricted regulations. We can allocate some pages of the web application for allowing access from the mobile application that responds to the web requests sent from the mobile http connection.

Because this approach of sending/receiving information is not encrypted, we use encryption algorithms known for both the web and mobile application. For extra security, we use a magic number used for hashing information that expires soon and need to be reconfigured by the web server.

The first attempt to connect to web server with the special code will gives the encrypted magic number to the mobile application. Then the user name and password will be sent using a hashing algorithm by the magic number as the salt, as the web *request* parameter. Then the server returns the authentication acknowledge back to the mobile application. Now every request from the server, such as request for updating and checking the status of the devices, must be accompanied by the hashed username and password. After some predefined time (e.g., 5 minutes), the magic number expires and the server data packets must include newly hashed username and password using the new magic number.

The above sequence is shown in Fig. 3, which also shows the sequence of main server application, retrieving information from the devices (here a robot) and database

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

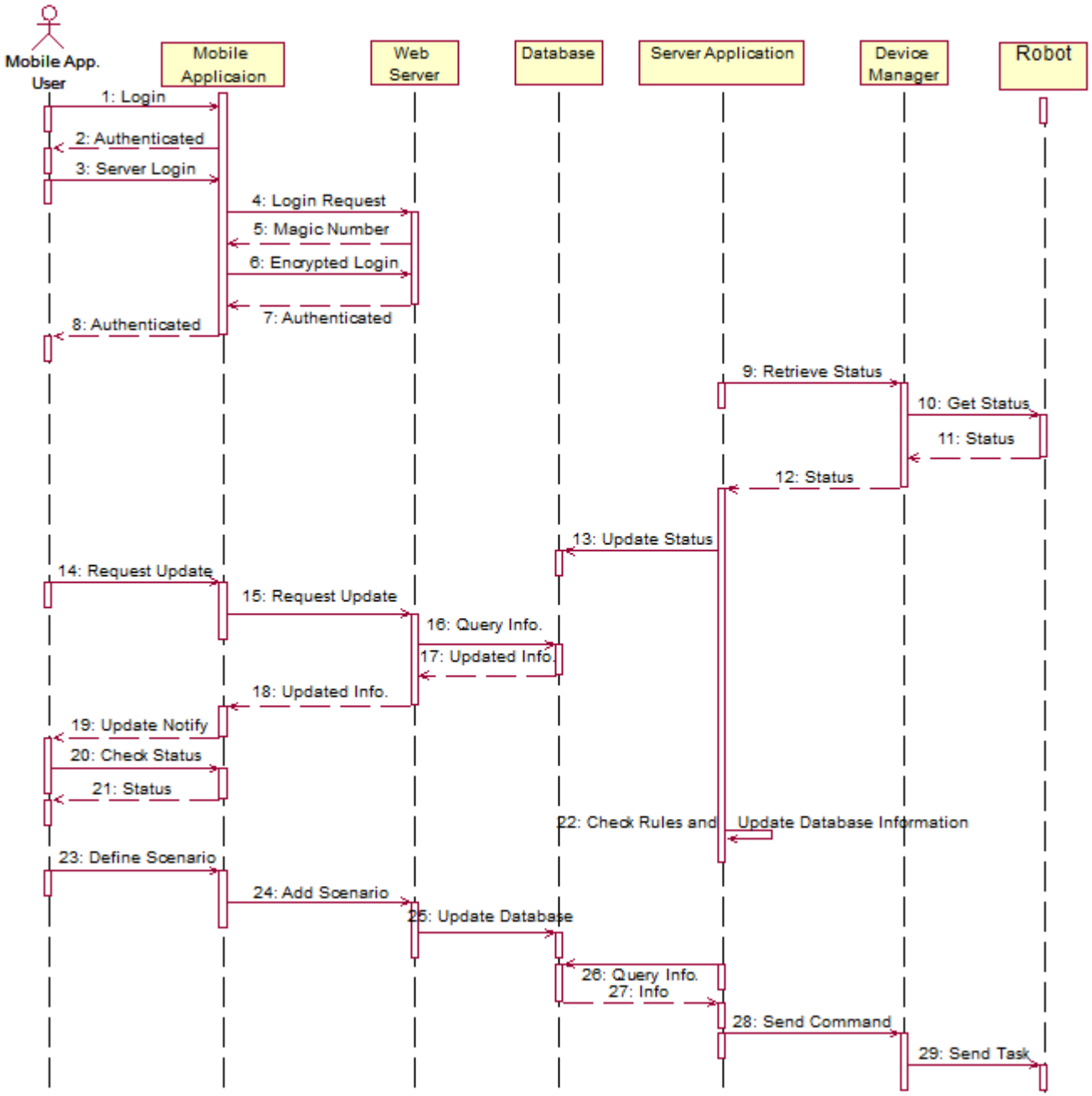


Figure 3. The sequence diagram of defining a robot-related scenario using the mobile application. This UML sequence diagram shows the necessary steps from logging into mobile application, defining the tasks and taking the appropriate actions to the robot as soon as application server becomes aware of it

while making the changes necessary. As soon as the mobile application or web application updates the database, the server will check the updated information, time and conditions, so that it will send necessary commands to the actuators to make the changes applied.

The database contains a table in which icons of the devices and their respective positions will be stored. This information is also transferred to the device in the case of updating the information. This table information is related but not depended on the main device table (in which detailed device controlling information is stored). In other way, the records and items in the home top view plane can be in the device records list either, but not necessarily all the map items must have full identification record in the device records table. This makes the whole house map items easy designing, but not limiting the selection to controllable devices. Now we must find a suitable way to transfer this map to the mobile device. The problem is the file size of the map that is too much to be transferred easily through the GPRS mobile internet; instead, we designed a way to create the map in the mobile device itself, using pre-defined icons.

IV. THE MOBILE APPLCATION IMPLEMENTATION

A. *Home Map*

A home map is necessary to simplify the access and status checking of the home devices and robots. Lu and Fu

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

proposed an activity map for a convenient and user-accessible interaction with the smart environment [12].

To transfer the data and create the map in the mobile device, first we must separate the map plane and objects within it. To do the task, as the device requests for updating the information of the devices and the maps, we send the house wall information as arrays of connected lines. To make this thing happen, we actually need a List<List<Points>> (a List of List of Points). The inner list contains points that makes continues lines within the points of the current list, making an open polygon designing possible. The first two points of each inner list will be used to determine the width and color of the lines. Because every point has two integer elements (x and y), two points make 4 integer data integration possible, so one is used for the line width and the other three represents an RGB color value.

For transferring the icons data, we used a record with the fields illustrated in Fig. 4. *OID* field, if not zero, can represent a device in the according device table in the database that can make the device selectable (e.g., for further details view and scenario/rule assignment). Other devices in the map have an *OID* of zero. *Name* and *Position* fields of the device map record are used for displaying purposes. The *IconID* field, regardless of the status of the device in the appropriate device table record, indicates the current status of the device/furniture using an icon in the mobile application icon database. The web page map controller is responsible for representing the appropriate *IconID* that best defines the *type* and *current status* of the device. For example, two icons can represent a door in two different statuses of being closed or opened.

OID	Name	Position	IconID

Figure 4. Fields in the record used to transfer home top view plane icons to the mobile application

For new devices that their icons is not available in the mobile application, some extra icons has been considered that makes the other unknown devices into 4 categories that can be recognized by their status easily:

1. On/Off devices (e.g., a lamp)
2. Leveled Devices (e.g., a gas sensor)
3. Appearing/Disappearing devices (e.g., a car or a bike)
4. Opened/Closed Devices (e.g., a door)

The device then repaints the map using the received information from the server, and the icons in its database. First the areas of the house are drawn using the line information defined by the points and the icons are painted just after it. The controllable devices in the map (whose OID is not zero) can be pointed and selected (like the application of the Gator Tech Smart House [3]) to check the status and define scenarios for it.

B. Retrieving Information from Server

For implementing the updating procedure, we divide the updating data into two categories. The first part is the devices data table which includes information about the devices, as well as the capabilities and controllable parts of each one. Because this information may be quite large to transfer and the devices and their controllable/sense-able features is device dependent, but not state dependent; they can be updated in longer periods than state information. This type of updating is labeled *Update Devices Data* in the main menu of the mobile application

The second part of the information is the device states and map information. Because this information is more likely to be updated, and contain less data than the first part, they can be downloaded every time the statuses are being checked. This regular information updating is accessible as *Update Information* button in the main menu form of the mobile application, as well as in *Check Status* and *Home Top Plane View* forms.

C. Windows Mobile platform Implementation

There are several platforms in which it is possible to implement the designed application, such as Java 2 Micro Edition (J2ME), Windows Mobile and Symbian. The J2ME is the most common platforms supported in mobile devices, but it's low level libraries makes it difficult to implement the application in the first place after designing it. So we decided to implement the application, in this stage, using Windows Mobile platform, which can help implementing the application with all the possible features as necessary.

1) Login Screens

For increased security, both the mobile application and server connection require a username and password. This was done so that the server address can be also protected from unauthorized viewing.

a) User Login

This screen simply contains the username and password fields for the user to access the application.

b) Server Login

Just like the simple User Login form, but with an additional textbox field labeled *Server path* that indicates the home web server address that the mobile application must connect to communicate with the server application.

2) Main Menu items

The main menu form of the application is the form appearing just after logging in. This form contains all the links necessary for different parts of the program (Fig. 5).

Update Information and *Update Devices data* was mentioned before. The *Home Top Plane View* will bring up the home top plane map of the house using the most recent updated information.

Live Camera Streaming is designed for live streaming the camera devices in the smart home. The *Manage Scenarios* and *Manage Rules* items are designed to list and manage the current device scenarios and conditional rules. Scenarios are set of tasks which will be done in a specific time (or current moment). Rules are condition and action sets in which the

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

condition are simultaneously checked and as soon as the criteria is met; the appropriate actions will be taken.

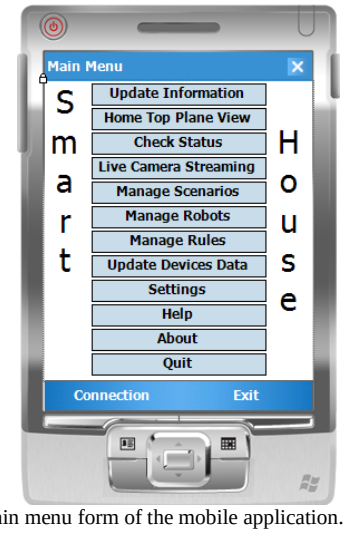


Figure 5. The main menu form of the mobile application. All different parts of the application can be accessed through this form.

Manage Robots will provide options regarding the robots, such as the list of all of them and the specific action each robot can do. The list is read-only in mobile application but the robots or their actions can be enabled/disabled for selecting in the scenario tasks.

The settings option will allow the user to change the server settings such as the address and update rules; as well as changing username/password and other settings.

3) *Home Map*

This form shows the home top plane map of the house, designed by the administrative user in the server computer, and downloaded as records of map elements by the mobile application (Fig. 6). If the information is out of date, it prompts the user to update the information. The user can also select the devices here and be prompted whether he/she wants to check its status or add a scenario for it and then will be redirected to the appropriate form. The icons in this map also represent the current status of the device as indicated by the *IconID* field of the device map record retrieved from the server.

Figure 6. Home top view plane of the house with selectable items.

4) *Managing Scenarios*

In this form, the updated scenarios will be listed and each of them can be modified and be enabled/disabled. New scenarios can also be created in a new form (Fig. 7). Each task has a name and will be applied on a device by the corresponding device action. The time of the scenario can be set to now, a specific time in the day or a specific time after execution time. Other defined scenarios can also be selected as a task in the new scenario. When a new task is selected to be designed, different robots can also be selected as the actor of the action and each action can take a value as the parameter. These data can be sent using SMS either.

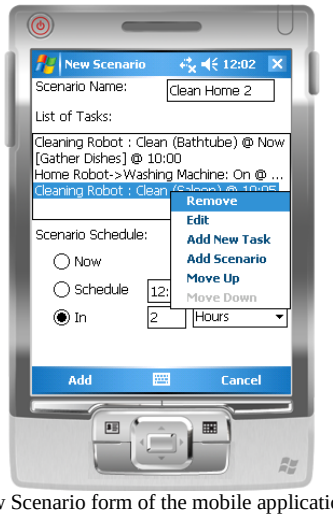


Figure 7. The New Scenario form of the mobile application. A list of tasks can be defined here including other scenarios or robot actions.

V. Future Works

As mentioned earlier, the works on a complete and comprehensive smart home that can work with all possible home appliances and can be controlled by all means in an effective way are all scattered around and researched independently. Only in some projects, some parts of a real smart home put into practice (e.g., The Gator Tech Smart House [13], and Plug and Play smart environment [14]). As we completed a comprehensive design from the previous work [15] and designed a server and mobile application needed for controlling the smart home remotely, we must continue the work on completing the whole server and applications, and move through commercial manufacturing of such houses so that all these efforts on designing the Smart Home can come to reality.

VI. Conclusion

In this paper, we presented an overview of a smart home control system server along with the way the devices are managed in the server. After discussing the possible security issues of developing a server for communicating to mobile application, we proposed a web server for the mobile application to communicate to it using GPRS. We presented the communication sequence through the web server in a UML sequence diagram and described the use-cases of both the server and mobile application. The scenarios were designed to set a number of tasks all in one place for further

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

and easier use. Robots participation and parameterized actions were also described along with regular actions. We finally explained the design of the mobile application and the data records needed for transferring the data and home top view plane from the server to mobile application in an efficient way. We finally described the main parts of the implementation of this smart home remote control mobile application in the Windows Mobile platform.

REFERENCES

[1] A. Sleman, M. Alafandi, and R. Moeller,"Integration of Wireless Fieldbus and Wired Fieldbus for Health Monitoring," Consumer Electronics, 2009. ICCE '09. Digest of Technical Papers International Conference, pp.1 – 2, 10-14 Jan. 2009

[2] T. Van Nguyen, J. Gook Kim, and D. Choi, "ISS: The Interactive Smart home Simulator," Advanced Communication Technology, 2009. ICACT 2009. 11th International Conference on , vol. 03, pp.1828-1833, 15-18 Feb. 2009

[3] C. Escoffier, J. Bourcier, P. Lalanda, and Y. Jianqi, "Towards a Home Application Server," Consumer Communications and Networking Conference, 2008. CCNC 2008. 5th IEEE, pp.321-325, 10-12 Jan. 2008

[4] Z. Salvador; R. Jimeno, A. Lafuente, M. Larrea, and J. Abascal, "Architecture for ubiquitous enviroments" Wireless And Mobile Computing, Networking And Communications, 2005. (WiMob'2005), IEEE International Conference on, pp.90-97, vol. 4, 22-24 Aug. 2005

[5] G. Patricio, and L. Gomes ,"Smart house monitoring and actuating system development using automatic code generation," Industrial Informatics, 2009. INDIN 2009. 7th IEEE International Conference on, pp.256-261, 23-26 June 2009

[6] H. Lee, Y Kim, K. Park, Z. Zenn Bien, "Development of a steward robot for human-friendly interaction," Computer Aided Control System Design, 2006 IEEE International Conference on Control Applications, 2006 IEEE International Symposium on Intelligent Control 2006 IEEE, pp.551 – 556, 4-6 Oct. 2006

[7] J. King, R. Bose, Y. Hen-I, S. Pickles, A. Helal,"Atlas: A Service-Oriented Sensor Platform: Hardware and Middleware to Enable Programmable Pervasive Spaces," Local Computer Networks, Proceedings 2006 31st IEEE Conference on, pp. 630 – 638, 14-16 Nov. 2006

[8] V. Santos, P. Bartolomeu, J. Fonseca, and A. Mota,"B-Live - A Home Automation System for Disabled and Elderly People," Industrial Embedded Systems, 2007. SIES '07. International Symposium on, pp.333-336, 4-6 July 2007

[9] H. Medjahed, D. Istrate, J. Boudy, and B. Dorizzi, "Human activities of daily living recognition using fuzzy logic for elderly home monitoring," Fuzzy Systems, 2009. FUZZ-IEEE 2009. IEEE International Conference on, pp.2001-2006, 20-24 Aug. 2009

[10] Yerrapragada, C.; Fisher, P.S.;" Voice Controlled Smart House"; Consumer Electronics, 1993. Digest of Technical Papers. ICCE., IEEE 1993 International Conference on, pp.154-155, 8-10 June 1993

[11] Sheng-Tzong Cheng; Chi-Hsuan Wang; Ching-Chung Chen; "An Adaptive Scenario Based Reasoning system cross smart houses" Communications and Information Technology, 2009. ISCIT 2009. 9th International Symposium on, pp.549 – 554, 28-30 Sept. 2009

[12] C. Lu, L. Fu, "Robust Location-Aware Activity Recognition Using Wireless Sensor Network in an Attentive Home," Automation Science and Engineering, IEEE Transactions on, vol. 6, Issue.4, pp.598-609, Oct. 2009

[13] S. Helal, W. Mann, H. El-Zabadani, J. King, Y. Kaddoura, and E. Jansen, "The Gator Tech Smart House: a programmable pervasive space", Computer, vol. 38, Issue. 3, pp.50 – 60, March 2005

[14] B. Abdulrazak, A. Helal, "Enabling a Plug-and-play integration of smart environments", Information and Communication Technologies, 2006. ICTTA '06. 2nd vol. 1, pp.820 – 825, 2006

[15] A. Rajabzadeh, A. R. Manashty, and Z. Forootan Jahromi, "A Mobile Application for Smart House Remote Control System," International Conference on Wireless Communication and Mobile Computing (ICWCMC 2010), Proceedings of WASET, vol. 62, ISSN. 2070-3724, pp. 80-86, February 2010

AUTHORS PROFILE

Amir Rajabzadeh received the B.S. degree in telecommunication engineering from Tehran University, Iran, in 1990 and received the M.S. and Ph.D. degrees in computer engineering from Sharif University of Technology, Iran, in 1999 and 2005, respectively. He is currently an assistant professor of Computer Engineering at Razi University, Kermanshah-Iran. He was the Head of Computer Engineering Department (2005–2008) and the Education and Research Director of Engineering Faculty (2008-2010) at Razi University.

Ali Reza Manashty is a senior B.S. student in Software Computer Engineering at Razi University, Kermanshah, Iran and will be graduated in September 2010 awarding the 3rd rank among the autumn 2006 entrance students of Computer Engineering Department. He is going to continue his academic career as a M.Sc. student in the next semester. He has been researching on mobile application design and smart environments especially smart digital houses since 2009. His publications include 4 papers in international journals and conferences and one national conference paper. He has earned several national and international awards regarding mobile applications developed by him or under his supervision and registered 4 national patents. He is a member of Exceptional Talented Students office of Razi University since 2008 and he was the teacher assistant of several under-graduate courses since 2008.

Zahra Forootan Jahromi is a senior B.S. student in Software Computer Engineering at Razi University, Kermanshah, Iran and will be graduated in September 2010. She is now researching in smart environments specially on simulating smart digital homes. Her publications include 4 papers in international journals and conferences and one national conference paper. She has 3 registered national patents and is now teaching Robocop robot designing for elementary and high school students at Alvand guidance school. She is a member of Exceptional Talented Students office of Razi University since 2008.

Predicting Students' Academic Performance Using Artificial Neural Networks: A Case Study

Ghaleb A. El-Refae
Faculty of Economics and Admin. Sciences
Al-Zaytoonah University of Jordan
Amman, Jordan
ghalebalrefae@yahoo.com

Qeethara Kadhim Al-Shayea
Faculty of Economics and Admin. Sciences, MIS Dep.
Al-Zaytoonah University of Jordan
Amman, Jordan
kit_alshayeh@yahoo.com

***Abstract*—Predicting students' academic performance is critical for universities because strategic programs can be planned in improving or maintaining students' performance. The goal of this study is to predict the factors affecting the university students' performance using Artificial Neural Networks (ANN) model. Various factors that may likely influence the performance of a student were identified. Generalized Regression Neural Network (GRNN) is used to predict the university students' performance. It is noticed a significant improvement in the prediction made by GRNN due to its generalization property. The most important predictor variable influencing performance is consistently having the largest regression. Results showed that secondary school performance which is measured by scores in secondary school certificate examination, measured in a percentage form having the largest regression value.**

Keywords-component; regression, stdudent performance; Artificial neural networks; general regression network

I. INTRODUCTION

The prediction and explanation of academic performance and the investigation of the factors relating to the academic success and persistence of students are topics of utmost importance in higher education [1].

McKenzie and Schweitzer [2] presented a study that was a prospective investigation of the academic, psychosocial, cognitive, and demographic predictors of academic performance of first year Australian university students.

Alfan [3] determined the undergraduate students' performance in the Faculty of Business and Accountancy, University of Malaya and the factors influencing the performance of the undergraduate students. The result of the study shows that the predictor variables do explain the variance in the students' final cumulative grade point average. In addition, it was found that knowledge prior to entering the university such as economics, mathematics and accounting is crucial in assisting the students in undertaking the courses in both business and accounting program. The study also found that female students perform better than male students; whilst Chinese students perform better than Malay and Indian students.

Su [4] evaluated the performance of university students who learned science texts by using, information communication technologies including animation, static figures, power point, and e-plus software. The results included the computation of the F-ratio, p-values, and Cohen's effect-sizes of attitudes toward science and learning science in relation to the student's gender, attendance of computer orientation classes, use of computer-multimedia, disposition toward computers, and majors.

Al-Tamimi and Al-Shayeb [5] investigated some factors affecting student performance in the fundamentals of financial management course at United Arab Emirates University.

Ibrahim and Rusli [6] developed three predictive models using SAS Enterprise Miner that are, artificial neural network, decision tree and linear regression. The result of this study showed that all of the three models produce more than 80% accuracy. It also showed that artificial neural network outperforms the other two models.

Oladokun, Adebanjo and Charles-Owaba [7] presented an artificial neural network model for predicting the likely performance of a candidate being considered for admission into the university was developed and tested.

II. ARTIFICIAL NEURAL NETWORKS

An artificial neural network (ANN) is a computational model that attempts to account for the parallel nature of the human brain. An (ANN) is a network of highly interconnecting processing elements (neurons) operating in parallel. These elements are inspired by biological nervous systems. As in nature, the connections between elements largely determine the network function. A subgroup of processing element is called a layer in the network. The first layer is the input layer and the last layer is the output layer. Between the input and output layer, there may be additional layer(s) of units, called hidden layer(s). Fig. 1 represents the typical neural network. You can train a neural network to perform a particular function by adjusting the values of the connections (weights) between elements.

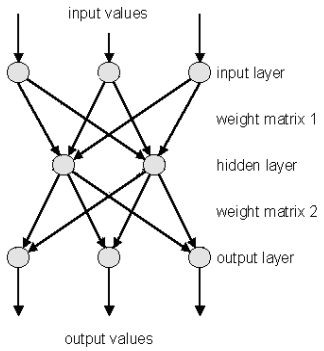


Figure 1. A typical neural network

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

For the researcher and the financial analyst, the main advantage of ANNs is that there is no need to specify the functional relation between variables. Since they are connectionist-learning machines, the knowledge is directly imbedded in a set of weights through the linking arcs among the processing nodes. In order to train a neural network properly one needs a large set of representative 'good quality' examples. In the case of bankruptcy problems, the researcher should be cautious when drawing conclusions from neural networks trained with only one or two hundred cases, as observed in most previous studies [8].

A. Generalized Regression Neural Network

The GRNN was applied to solve a variety of problems like prediction, control, plant process modeling or general mapping problems [9].

General regression neural network Specht [10], Nadaraya [11] and Watson [12], does not require an iterative training procedure as in back-propagation method.

The GRNN is used for estimation of continuous variables, as in standard regression techniques. It is related to the radial basis function network and is based on a standard statistical technique called kernel regression. By definition, the regression of a dependent variable y on an independent x estimates the most probable value for y, given x and a training set. The regression method will produce the estimated value of y, which minimizes the mean-squared error. GRNN is a method for estimating the joint probability density function (pdf) of x and y, given only a training set. Because the pdf is derived from the data with no preconceptions about its form, the system is perfectly general. Furthermore, it is consistent; that is, as the training set size becomes large, the estimation error approaches zero, with only mild restrictions on the function. In GRNN, instead of training the weights, one simply assigns to wij the target value directly from the training set associated with input training vector i and component j of its corresponding output vector [13]. GRNN architecture is given in Fig. 2. GRNN is based on the following formula [14]:

$$E[y \mid x] = \frac{\int_{-\infty}^{\infty} y.f(x,y).dy}{\int_{-\infty}^{\infty} f(x,y).dy} \quad (1)$$

where y is the output of the estimator, x is the estimator input vector, $E[y|x]$ is the expected output value, given the input vector x and $f(x,y)$ is the joint probability density function (pdf) of x and y.

The function value is estimated optimally as follows:

$$y_j = \frac{\sum_{i=1}^{n} h_i.w_{ij}}{\sum_{i-1}^{n} h_i} \quad (2)$$

where w_{ij}= the target output corresponding to input training vector x_i,

$h_i = e^{\frac{-D_i^2}{2.spread^2}}$, the output of the hidden layer neuron,

$D_i^2 = (x-u_i)^T (x-u_i)$, the squared distance between the input vector x and the training vector u, x= the input vector, u_i=training vector i, the center of neuron i, spread=a constant controlling the size of the receptive region.

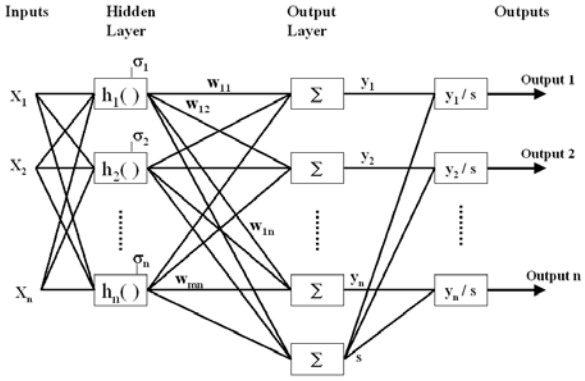


Figure 2. Generalized Regression Neural Network (GRNN) Architecture

III. Experimental Results

A. Data

This study was conducted at the faculty of Economics and Administrative Sciences, Al-Zaytoonah University of Jordan in Hashemite Kingdom of Jordan. Our sample consists of 208 students belonging to accounting department. The information for this study has been obtained from the register office at Al-Zaytoonah University of Jordan and which are maintained on a computerized database.

The Cumulative Grade Average Point (CGPA) is used as an indicator to measure the performance of the university students'.

The students' overall performance was hypothesized to be a function of the following factors: (1) Secondary school performance is measured by scores in secondary school certificate examination, measured in a percentage form (2) type of secondary school branch, (3) gender, and (4) boarding or non boarding student.

B. Results Analysis

A generalized regression neural network (GRNN) with a radial basis layer and a special linear layer and linear output neurons was created using the neural network toolbox from Matlab 7.9 as shown in Fig. 2. Generalized regression neural networks are a kind of radial basis network that is often used for function approximation.

The first layer has as many neurons as there are input/ target vectors. Each neuron's weighted input is the distance between the input vector and its weight vector. Each neuron's net input is the product of its weighted input with its bias. Each neuron's output is its net input passed through radial basis transfer function. Radial basis transfer function is a neural transfer function which calculates a layer's output from its net input. If a neuron's weight vector is equal to the input vector (transposed), its weighted input will be 0, its net

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

input will be 0, and its output will be 1. The second layer also has as many neurons as input/target vectors.

We used a spread slightly lower than the distance between input values, in order, to get a function that fits individual data points fairly closely. A smaller spread would fit data better but be less smooth.

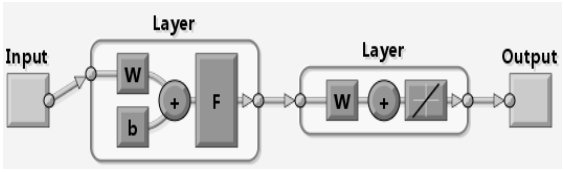


Figure 3. A generalized regression neural network (GRNN)

The GRNN with cumulative grade point average CGPA as a target and secondary school performances that is measured by scores in secondary school certificate examination, measured in a percentage form as input was been created. Then simulate the network with 208 inputs. The network outputs after simulation. The spread value was chosen 0.2. The percent correctly predicted in the simulation sample is approximately 76 percent as shown in Fig. 4.

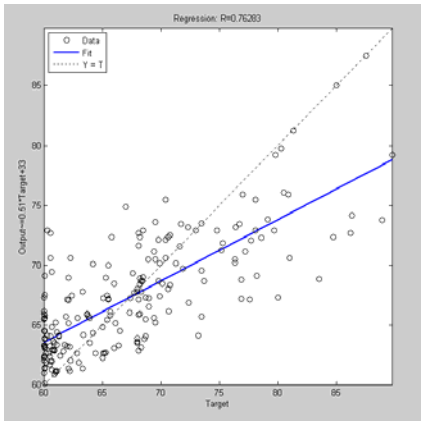


Figure 4.

The GRNN with cumulative grade point average CGPA as a target and type of secondary school branch as input was been created. The spread value was chosen 0.6. The percent correctly predicted in the simulation sample is approximately 36 percent as shown in Fig. 5.

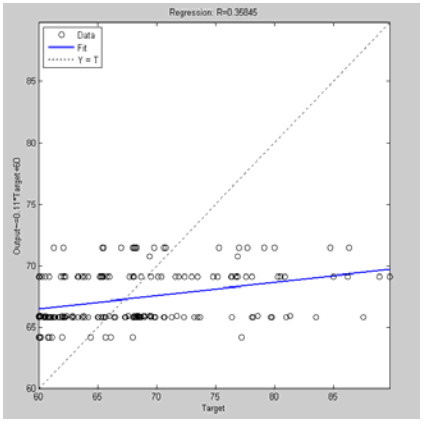


Figure 5.

The GRNN with cumulative grade point average CGPA as a target and gender as input was been created. The spread value was chosen 0.5. The percent correctly predicted in the simulation sample is approximately 27 percent as shown in Fig. 6.

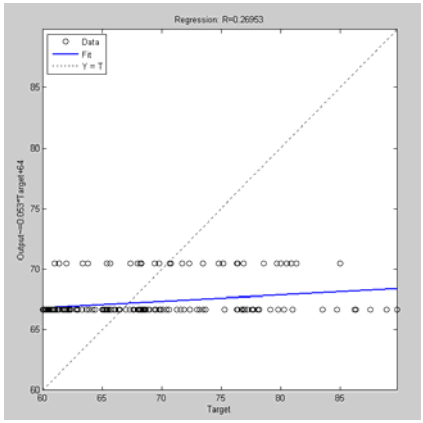


Figure 6.

The GRNN with cumulative grade point average CGPA as a target and boarding or non boarding as input was been created. The spread value was chosen 0.5. The percent correctly predicted in the simulation sample is approximately 20 percent as shown in Fig. 7.

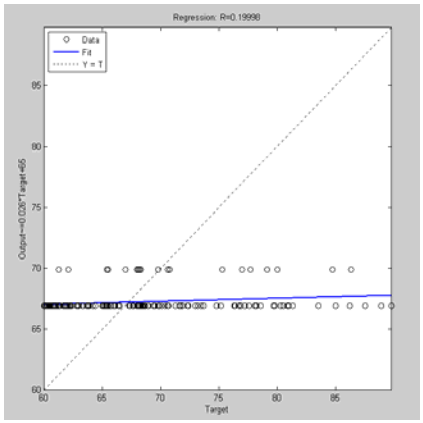


Figure 7.

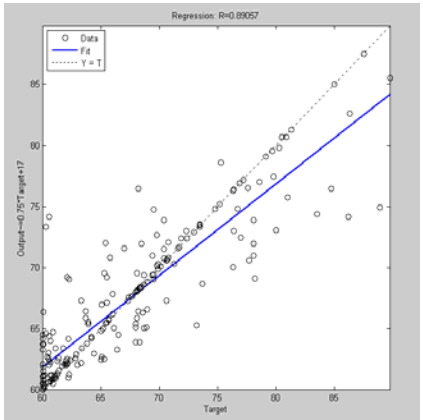


Figure 8.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

It is obvious from the results, that the most affecting factor in academic university students' which is measured by CGPA is the secondary school performance.

Fig. 8 shows the multiple regressions for the four affecting factors. The spread value was chosen 0.2. The percent correctly predicted in the simulation sample is approximately 89 percent.

IV. CONCLUSIONS

In this paper the general regression neural network is used for the prediction of university student performance. The advantage of using the GRNN in the prediction is its generalization property. The results of this study provide evidence which suggests that secondary school performance is the single most important variable associated with their overall performance upon graduation from university. Other variables such as type of secondary school branch, gender, and boarding or non boarding student show a lesser degree of significance in predicting performance as compared with secondary school score. The results also indicate that the variables examined in this study provided a significant contribution in predicting performance when used jointly with secondary school performance variable.

References

[1] P. Fenollar, S. Roma´n and P. J. Cuestas, University students' academic performance: An integrative conceptual framework and empirical analysis, British Journal of Educational Psychology, 77, pp. 873–891, 2007.
[2] K. Mckenzie and R. Schweitzer, Who Succeeds at University? Factors Predicting academic performance in first year Australian university students, Higher Education Research and Development, Vol. 20, Issue 1, pp. 21-33, May, 2001.
[3] E. Alfan, Undergratuate Students' Performance: The Case of University of Malaya, Quality Assurance in Education: An International Perspective, Vol. 13, Issue 4, pp. 329-343, 2005.
[4] K. Su, An integrated science course designed with information communication technologies to enhance university students' learning performance, Computers and Education, Vol. 51, Issue 3, pp. 1365-1374, Nov., 2008.
[5] H. A. Al-Tamimi and A. R. Al-Shayeb, Factors Affecting Student Performance in the Introductory Finance Course, Journal of Economics and Administrative Science, Vol. 18, No. 2 Dec., 2002.
[6] Z. Ibrahim and D. Rusli, Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision Tree and Linear Regression, 21st Annual SAS Malaysia Forum, 5th September 2007.
[7] V.O. Oladokun, A.T. Adebanjo, and O.E. Charles-Owaba, Predicting Students' Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course, The Pacific Journal of Science and Technology, Vol. 9, No. 1, pp. 72-79, May, 2008, web page available at:
[8] J. C. Neves and A. Vieira, Improving Bankruptcy Prediction with Hidden Layer Learning Vector Quantization, European Accounting Review, Vol. 15, No. 2, pp. 253–271, 2006.
[9] D. W. Patterson, Artificial Neural Networks, Theory, and Applications, Englewood Cliffs, NJ: Prentice-Hall, 1995.
[10] D. F. Specht, A general regression neural network, IEEE Trans. Neural Networks, vol. 2, pp. 568–576, 1991.
[11] E. A. Nadaraya, On estimating regression, Theory of Probab. Applicat., vol. 9, pp. 141–142, 1964.
[12] G. S. Watson, Smooth regression analysis, Sankhya Series A, vol. 26, pp. 359–372, 1964.
[13] M. T. Hagan, H. B. Demuth, M. Beale, Neural network design, PWS Publishing Company, Boston, 1996.
[14] K. Kayaer and T. Yildirim, Medical Diagnosis on Pima Indian Diabetes Using General Regression Neural Networks, web page available at: www.yildiz.edu.tr/~tulay/publications/Icann-Iconip2003-2.pdf.

AUTHORS PROFILE

Ghaleb A. El-Refae, has a Ph. D. and M.A in Financial Economics form USA, M. Sc and B. Sc in Accounting. He is a professor and Dean of Faculty of Economics and Administrative Sciences at Al-Zaytoonah University of Jordan, Amman – Jordan. His research interest is in the application of IT and IS in Business and Economics.

Qeethara Kadhim Abdul Rahman Al-Shayea, has received Ph. D. in Computer Science, Computer Science Department, University of Technology, Iraq, 2005. She received her M.Sc. degree in Computer Science, Computer Science Department from University of Technology, Iraq, 2000. She has received her High Diploma degree in information Security from Computer Science Department, University of Technology, Iraq, 1997. She has received B. Sc. Degree in Computer Science Department from University of Technology, Iraq, 1992. She joined in September (2001-2006), Computer Science Department, University of Technology, Iraq as assistant professor. She joined in September 2006, Department of Management Information Systems Faculty of Economics & Administrative Sciences Al-Zaytoonah University of Jordan as assistant professor. She is interested in Artificial intelligent, image processing, computer vision, coding theory and information security.

Vast Educational Mobile Content Broadcasting using ARMrayan Multimedia Mobile CMS

Somayeh Fatahi
Department of Computer Engineering
Kermanshah University of Technology
Kermanshah, Iran
fatahi_somayeh@yahoo.com

Ali Reza Manashty and Zahra Forootan Jahromi
Department of Computer Engineering
Razi University
Kermanshah, Iran
a.r.manashty@gmail.com, zahra.forootan@gmail.com

***Abstract*— The huge information flow currently available to young generation makes it difficult for educational centers to train them as needed. Most of these information flows occur in transportation time or while on public areas. Competing with commercial information streams is far out of educational centers time and budget. For creating enough mobile applications for vast educational mobile content broadcasting that can match young spirits as well, we designed and developed the ARMrayan Multimedia Mobile CMS as the software that helps communities, educational, cultural or marketing centers in a way that ordinary operators be able to create a variety of fully functional multimedia mobile applications such as tutorials, catalogues, books, and guides in minutes without writing even a line of code. In this paper, we present the role of our developed software in our proposed vast educational content broadcasting system using kiosks and Bluetooth advertising, which will lead to a great leap in M-commerce marketing and public education solutions. Related experiences are described and diagrams are used to illustrate the solution. Upon release of the software, it achieved two titles and prizes in different festivals and various cultural and commercial centers became its customers.**

Keywords- mobile education; m-commerce; mobile CMS; multimedia cms; mobile; content broadcasting; mobile catalogue; education; J2ME.

I. INTRODUCTION

Nowadays, the use of cell phones as an everywhere company has become a habit for nearly everyone. E-Commerce, which is based on using electronic devices for business uses electronic marketing as the main means of advertising. While e-commerce mostly relies on web and internet, an area which focuses mostly on marketing and trading using the mobile phones of the consumers, is called Mobile Commerce (M-Commerce). This area of commerce mostly involves contacting users via their mobile phones. Recent business models and applications of M-Commerce are fully described by Chen Xin [1]. This includes using mobile networks to send text/multimedia contents such as SMS and MMS to proximity marketing which is based on sending mobile contents using Bluetooth technology to the cell phones in the range of the Bluetooth sender device.

The huge information flow currently available to young generation makes it difficult for educational centers to train them as needed. Most of these information flows occur in transportation time or while on public areas. Competing with commercial information streams is far out of educational centers time and budget. For creating enough mobile applications for vast educational mobile content broadcasting that can match young spirits as well, we designed and developed the ARMrayan Multimedia Mobile CMS as the software that helps communities, educational, cultural or marketing centers in a way that ordinary operators be able to create a variety of fully functional multimedia mobile applications such as tutorials, catalogues, books, and guides in minutes without writing even a line of code.

Public education affects many areas. First it is closely in contact with industry as both affect each other in budget and policy [2]. In long term, public education also helps removing the inequality between poor and rich [3] and help communities improve education quality [4], so the best practice for improving justice in society education is investing in innovative public education methods.

In proximity marketing, as mentioned earlier, the contents that can be sent using the Bluetooth devices vary from virtual cards (VC) to text and multimedia contents such as sound, video and picture and even mobile applications. It is nearly impossible to send various different contents such as a text and picture using for example a single HTML file, because most of the users only accept the incoming connection once from each Bluetooth advertising devices and most possibly will be out of range of the device (lose connection) after the first file received. The later problem is a permanent problem because of the limited range of Bluetooth 2.0 EDR technology for mobile phones which is at most about 150 meters.

The problem of sending various related contents using proximity marketing was a challenge which was resolved with the use of mobile E-book makers, especially those with the capability of displaying Unicode language of the region regardless of the target mobile device support. The output of such mobile E-book makers is a stand-alone J2ME mobile application which can be sent as a single .jar file and is

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

installable on most of the mobile phones that support Java platform.

Most of the primary mobile e-book makers, especially in Iran, which could support Farsi language, such as Parnian Mobile E-book Studio, could only support text and a theme. The next generation of these E-book makers could support pictures in the text too. Still none of them has the capability of supporting rich media contents such as sounds and videos.

Nowadays people spend a notable amount of time in transportation, without having access to their PCs or having hard time accessing their laptops; instead, they are constantly using their cell-phones/PDAs. Because of this, many companies and active advertisement centers, whether religious, educational or business related are pinpointing on these handset devices for applying their policies. Despite of this huge trend, there is still a lack of programmers that write both visually and functionally acceptable mobile applications in beneficial time and cost.

Due to the lack of rich media support in such content-oriented mobile applications and the need for fast production of such mobile applications, we designed and implemented ARMrayan Mobile Multimedia CMS (Content Management System) [5] to omit the programming part of producing content oriented multimedia mobile software. This software help communities, educational, cultural and marketing centers in a way that ordinary operators be able to create a variety of fully functional multimedia mobile applications such as tutorials, catalogues, books, and guides in minutes without writing even a line of code.

In this paper, we present the role of our developed software in vast educational proximity marketing systems using Bluetooth advertising, which will lead to a great leap in M-commerce marketing solutions. Educational and cultural rich content mobile applications can now be produced easily and distributed using proximity marketing technologies, reaching the goal for a higher educated and cultural society. Related experiences are described and diagrams are used to illustrate the solution that involves a two level updating procedure for kiosks that involve end-users using proximity marketing features.

II. PROPOSED METHOD

As mentioned earlier, people waste a lot of time in transportation. In big cities, this transportation might take more than an hour. There has been many commercial products which involve taking advantages of these wasted times of people with teaching skills they mostly need to learn, such as language and psychological skills. Still there is lack of governmental effort to take advantage of these wasted times. Non-electronic cultural media such as small books and newspaper is not that suitable because of their volume and free-space consumption. A heavy bulk of newspapers and books will tire even the strongest people. Here is where the important point emerges: Mobile phones that are carried by all people every day.

A. Importance

By end-2008, there were nearly 3.9 billion mobile subscribers worldwide and it is expected that in 2013 this number will reach close to 6 billion [6]. The fact above, shows that even people are not eager to carry books and paper advertisement everywhere, mostly in transportation, they will not leave behind their mobile phones. Users' mobile phones are the greatest nest for cultural contents and commercial ads. With operators mobilizing to explore and capture potential markets in the developing nations, it is expected that Asia Pacific will house nearly 50 percent of the worldwide mobile subscriber base by end-2013 [6]. This also means that the continent of culture and tradition will be the host of half of the cell phone carriers in near future. Governments of Asian countries are the most possible demanders of transferring their traditional to the new generation to keep it more alive than ever.

B. ARMrayan Multimedia Mobile CMS

Now that mobile devices are non-detachable part of everybody's routine; there is still a need for mobile content creation tools so that the providers of contents can publish their data (whether rich-content data or simple texts) through structured theme-based applications that can also provide some facilities for using mobile device special capabilities (e.g., sending contents through SMS/MMS or media playback; that using mobile browsers for these jobs is still facing some difficulties).

In web-based services, the term CMS (Content Management System) is used to describe a web application that gives content-oriented accessibility to the administrator of the website, regardless of his/her knowledge in programming concepts of the website. In addition to content-oriented access, the administrator can also change the visual theme of his/her website. There's still a need for a system that provides tools for easy management of contents for use in mobile application providing additional mobile device facilities for efficient distributing or viewing of mobile-based contents. Such applications may now be called Mobile Content Management Systems or MCMSs.

The ARMrayan Multimedia MCMS is a PC-based application providing the user the tools needed for inserting his/her desired multimedia content in a tree-based index of pages, in addition to providing simple theme-based UI designing facilities. The role of our implemented software is to simplify the content transferring process by creating a means for producing rich media content-oriented mobile software. The output of this Windows application is a J2ME mobile application with the desired content and visual design. The J2ME application was developed separately using the Java language for MIDP2.0+ supporting mobile phones. This base J2ME file that the Windows program receives as input is called the Mobile Engine File. This Engine File can be modified or updated regularly to maintain compatibility with newer devices and/or get improvement in the UI or search engine features.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

1) Implementation and Usage

For using ARMrayan MCMS, an operator, managed by the system administrator, designs a tree index of pages in the Windows program and then he/she will now be able to add several contents per page in the desired order. These content types can currently be among texts, pictures, sounds or music, videos and animations, map points, phone number, e-mail address, web sites link and etc. It is also possible to change the order of the contents within each page. The colors, background pictures and the background music of different areas of the output mobile application can also be set in the main user interface.

Implementation of the Windows application was using Visual C# .Net while the mobile engine application was developed in Java 2 Micro Edition language. The main features of the implemented software are as following:

- Supporting texts, sounds, images, videos, phone numbers, web links and emails as the input content.
- Tree-view page designing.
- Displaying Middle Eastern fonts via the bitmap font framework.
- Multilanguage environment (English and Farsi)
- Environment theme designing.
- Sending contents via SMS/MMS.
- Text searching.

2) Preview

The main program engine displays the content in tree views thus making it possible for a page to have innumerous subpages (Fig. 1). The mobile engine of the ARMrayan MCMS was developed as J2ME application. This engine is actually a stand-alone mobile application simply known as a "Mobile Engine" in the application, containing the main algorithms for displaying or playing different contents such as sounds, videos and texts. This mobile software (Fig. 2) also contains the functions to show the menus and performing other operations such as searching within the content and handling the bitmap font for Asian languages.

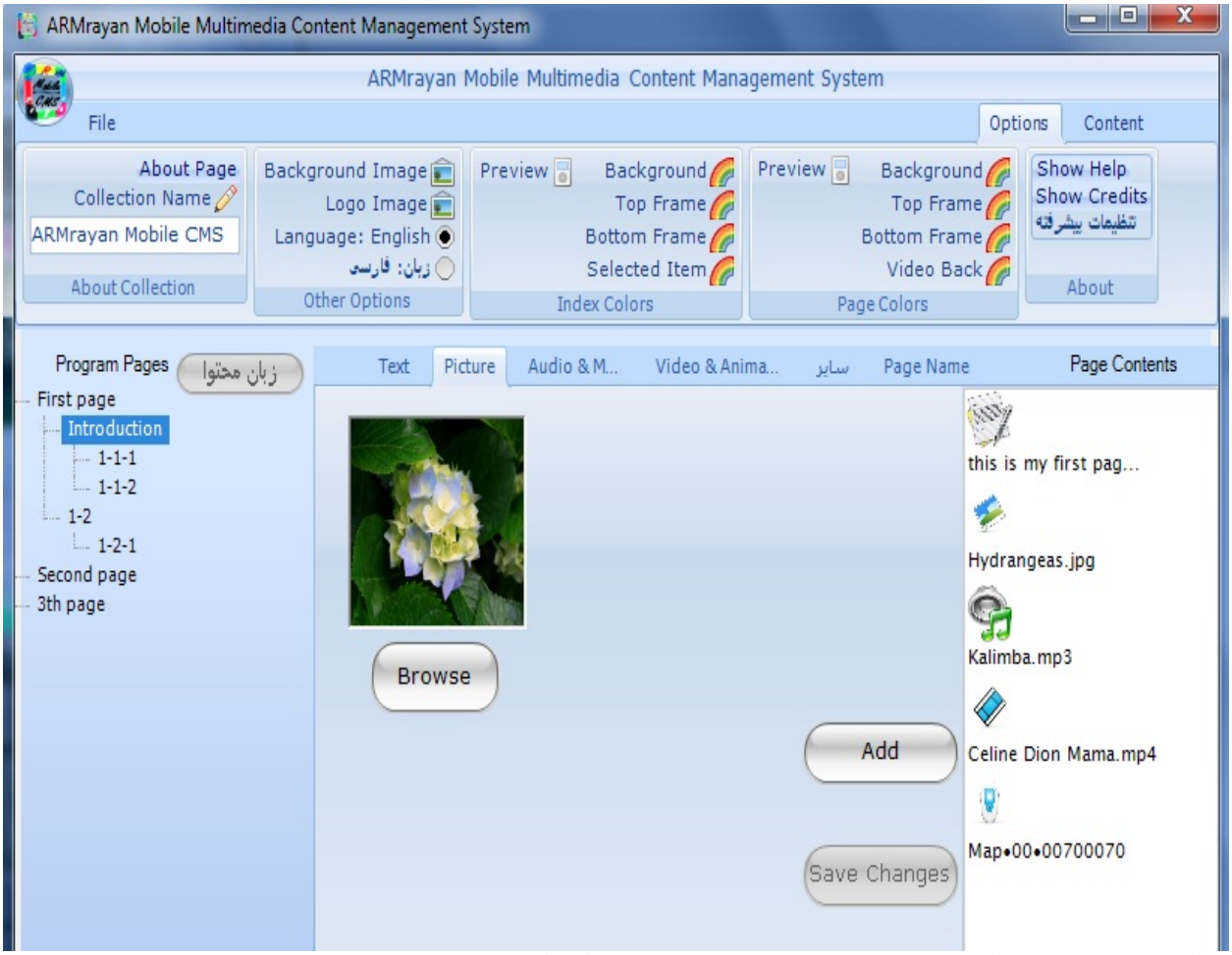


Figure 1. ARMrayan Multimedia Mobile CMS, Windows interface for inserting multimedia contents and designing final mobile application for creating final mobile application

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Figure 2. The output Java mobile application resulted from the design and contents of ARMrayan MCMS illustrated in (Fig. 1).

C. Vast Educational Mobile Content Broadcasting

It was described earlier that the most vital part of each society that can be the within the range of electronic cultural and educational goods, are the people using cell phones every day, mostly in transportations. This can be a target for both educational and cultural centers to broadcast their content to the society. This is done by gathering contents from certified sources and encapsulating the contents in a well-designed mobile application. Now we describe our proposed system for broadcasting vast educational multimedia mobile contents that is produced using ARMrayan Multimedia Mobile CMS and will be delivered using Proximity Marketing Systems.

1) Creating and Broadcasting Contents

An operator, under the supervision of an administrator, inserts multimedia contents to the ARMrayan Multimedia MCMS and receives a Java mobile application (a .jar file) as the output. While categorizing the content-based applications for broadcasting in several servers, the operator uploads the file to a central server. Various other sub-servers collect their appropriate categorized applications, adding new ones or updating current mobile applications, and start updating the several Proximity Marketing Systems that are in contact with the end-users (people) (Fig.3).

2) Sub-Servers

The various sub-servers that are designed to broadcast a series of categorized applications can be connected to multiple Proximity Marketing Systems (such as kiosks) in different places of a metropolitan area. After being updated from the main server, the sub-servers can disconnect from the main server and start updating the kiosks under their control. The category and kiosk locations can vary by the end-user groups and content intended to broadcast. For example a commercial advertisement group may use a dedicated sub-server to broadcast advertisement contents in various large malls and shopping areas. On the other hand, a university can publish its latest discoveries and research results along with educational contents and tests using a server while other governmental cultural centers can place such kiosks in public transportation areas where people can automatically or manually receive educational mobile e-books with rich multimedia contents that are easily readable and portable within transportation stations and along the day.

3) Proximity Marketing Systems

The Proximity Marketing Systems used in this project are some kiosks that have embedded Bluetooth devices that can function in two different ways:

- Automatically broadcasting mobile application software that is designed to be shared to general public using Bluetooth technology within the range of the device.
- Manually receivable mobile applications that is chosen from the inner menu of the kiosk stations. This offers a wider variety of mobile applications to be shared for specific persons that intend to find their suitable content themselves.

III. Conclusion and Results

We've tested a Proximity Marketing System for one of our mobile application files in the 3rd International Digital Media Festival and Exhibition, Tehran, October 2009. There was one server that was automatically sending the file to the customers. In each Bluetooth search the device was making, it could discover about 180 devices but most of them were vanishing too fast. The device could only send 7 files simultaneously within the range of about 100 meters in radius. In a 2-day test, a total of about 1800 attempt to send the file has been made by the device, from which about 600 were successful, 200 failed and 1000 were rejected by the customers. During this time, no more than about 100 visits were made to our place, where some of which found us using

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

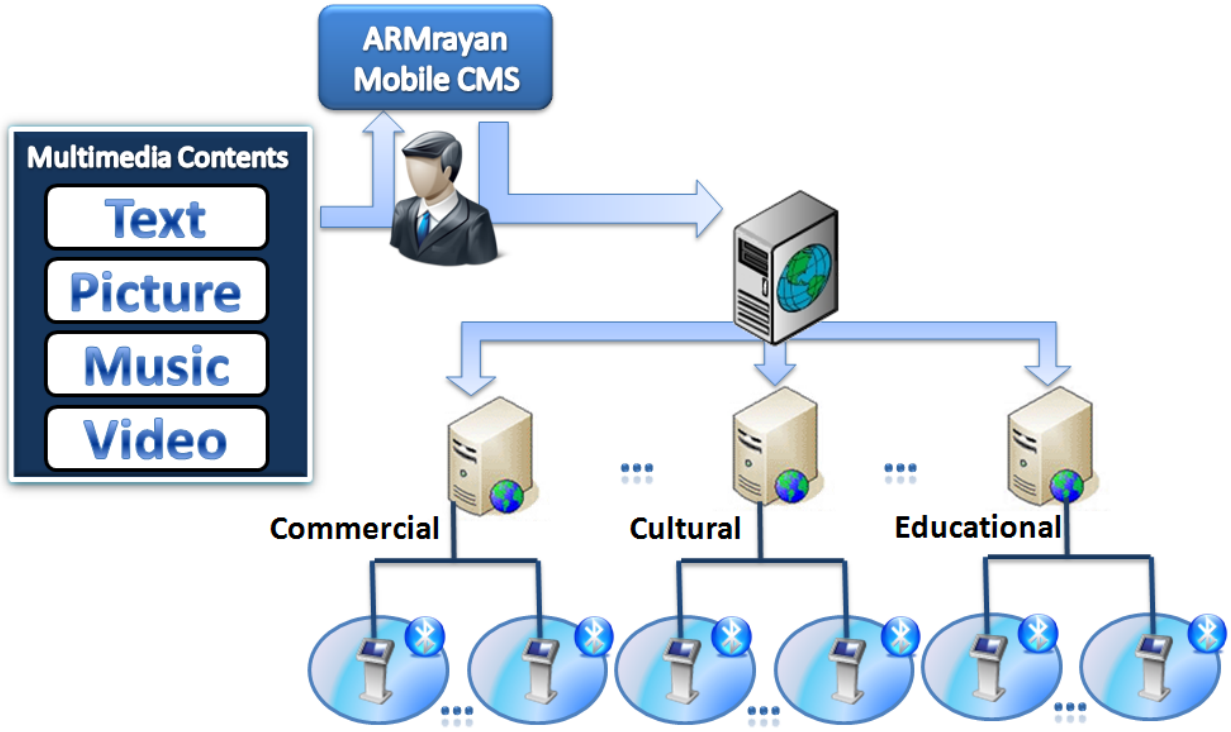


Figure 3. The proposed system diagram of Public Education and M-Commerce via Proximity Marketing Systems using ARMrayan Multimedia Mobile CMS.

the file we sent them earlier! It can be inferred from this attempt that:

- About 6 times more than the people that physically visited the commercial place, the people received the advertising information electronically using the Proximity Marketing System.
- More than 50% of the users rejected the file transfer, which means there need to be a huge effort for creating a safe and informative atmosphere so that people accept such files more easily and safely.

Such an experience repeated with just the same results, but with a decrease in rejections in the tourism activities in Norouz 89 (Starting of the New Year in spring). The reason of the reduced rejects was the banners informing the tourists to turn their Bluetooth on and accept the incoming file to inform them better about the travel they would made.

The multi-server based project described in this paper was proposed to the Kermanshah City government to be executed in several public transportation stations.

In the first month of developing this system, we awarded the 1st Place in the 3rd International Digital Media Festival as the "The Best Mobile Software in Technology, Innovation and Development" for "ARMrayan Mobile CMS"; Iran-Tehran; October 2009 [7]. Not many days passed, that an output of the program which contained audio and photo gallery, multi-language text and fabulous design, also awarded the 3rd Place in the 4th National Imam-Reza Festival, Professional Mobile Software Title of the Digital Media section, for the "Imam Reza Pilgrim Mobile Software"; Iran-Tehran; December 2009 [8]. We also sponsored by a high cultural and religion center for upgrading the application with their desired need, so that they can produce multi-language cultural and religious content-oriented mobile application for broadcasting into the society.

IV. Future Works

Our software is Windows-based and operator should upload the mobile application output file to the server upon creating it. It is considered to design and implement the ARMrayan Multimedia Mobile CMS as a web application in near future, so that the operator can create specific mobile application directly on the web by using the web based version of the software.

The other work that can be done in future is finding a way that users of our mobile applications are able to update their applications, which have installed on their phone, via using GPRS capability of mobile phones and get the new version of our different mobile applications. This is a very good solution for our users to be able to update their installed mobile applications without referring to the kiosks which are installed in a location that they might not be able to access it directly. A similar research but in another area of M-Commerce was done by Fumin Zou Shuling Zhang [9], where he presented a model for internet system for M-Commerce in trains.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Although some research has been done about the factors affecting people in mobile advertising from mobile operators [10, 11, 12], more research on the following area is considered helpful, with consideration of people's opinion in several places:

- What are the effects of the banners on people's acceptance or rejection of the file?
- What type of file do people mostly prefer to accept?
- What is the effective area of each Proximity Marketing System?
- What is the efficient size of the file to send to the people?
- Receiving the feedback of people about the contents they received and the one they desire the most.

REFERENCES

[1] C. Xin, "Business Models and Application of M-Commerce," Intelligent Information Technology Application Workshops, 2009. IITAW '09. Third International Symposium on, 21-22 Nov. 2009, pp. 98-100, 21-22 Nov. 2009

[2] K. Sano and Y. Tomoda, "Optimal public education policy in a two sector model," Economic Modelling, vol. 27, Issue. 5, pp. 991-995, September 2010

[3] G. Glomm and B. Ravikumar, "Public education and income inequality," European Journal of Political Economy, vol. 19, Issue. 2, pp. 289-300, June 2003

[4] M. J. Luengo-Prado, O. Volij, "Public education, communities and vouchers," The Quarterly Review of Economics and Finance, vol. 43, Issue. 1, pp. 51-73, 2003

[5] A. R. Manashty, M. R. Ahmadzadeh Raji, Z. Forootan Jahromi, and A. Rajabzadeh, "ARMrayan Multimedia Mobile CMS: a Simplified Approach towards Content-Oriented Mobile Application Designing," International Conference on Wireless Communication and Mobile Computing (ICWCMC 2010), Proceedings of WASET, vol. 62, ISSN. 2070-3724, pp. 62-67, February 2010

[6] http://www.portioresearch.com

[7] http://www.DMF.ir/

[8] http://www.imamreza-tehran.ir/

[9] F. Z. S. Zhang, "An Open Onboard Internet System Model for the M-Commerce on Train", Management and Service Science, 2009. MASS '09. International Conference on, pp. 1–5, 20-22 Sept. 2009

[10] S. Soroa-Koury and K. C.C. Yang, "Factors affecting consumers' responses to mobile advertising from a social norm theoretical perspective," Telematics and Informatics, vol. 27, Issue. 1, ISSN. 0736-5853, pp. 103-113, February 2010

[11] E. B. Cleff, "Effective approaches to regulate mobile advertising: Moving towards a coordinated legal, self-regulatory and technical response," Computer Law & Security Review, vol. 26, Issue 2, pp.158-169, ISSN. 0267-3649, March 2010

[12] E. B. Cleff, "Regulating mobile advertising in the European Union and the United States, Computer Law & Security Report," vo. 24, Issue. 5, ISSN 0267-3649, pp. 421-436, 2008

AUTHORS PROFILE

Somayeh Fatahi received her B.Sc. and M.Sc. degrees in Computer Engineering from Razi University, and Isfahan University, Iran, in 2006 and 2008 respectively and is now a Ph.D. student in University of Tehran, Iran. She is Lecturer in the Department of Computer Engineering at Kermanshah University of Technology. She also teaches in Razi University, University of Applied Science and Technology, Payam Noor University of Kermanshah, Islamic Azad University of Kermanshah, Institute of Higher Education of Kermanshah Jahad-Daneshgahi. Her research activities include (1) simulation of agents with dynamic personality and emotions (2) Computational cognitive modeling (3) Simulation and formalization of cognitive processes (4) Multi Agent Systems (5) Modeling of Human Behavior (6) Fuzzy Expert Systems.

Ali Reza Manashty is a senior B.S. student in Software Computer Engineering at Razi University, Kermanshah, Iran and will be graduated in September 2010 awarding the 3rd rank among the autumn 2006 entrance students of Computer Engineering Department. He is going to continue his academic career as a M.Sc. student in the next semester. He has been researching on mobile application design and smart environments especially smart digital houses since 2009. His publications include 4 papers in international journals and conferences and one national conference paper. He has earned several national and international awards regarding mobile applications developed by him or under his supervision and registered 4 national patents. He is a member of Exceptional Talented Students office of Razi University since 2008 and he was the teacher assistant of several under-graduate courses since 2008.

Zahra Forootan Jahromi is a senior B.S. student in Software Computer Engineering at Razi University, Kermanshah, Iran and will be graduated in September 2010. She is now researching in smart environments specially on simulating smart digital homes. Her publications include 4 papers in international journals and conferences and one national conference paper. She has 3 registered national patents and is now teaching Robocop robot designing for elementary and high school students at Alvand guidance school. She is a member of Exceptional Talented Students office of Razi University since 2008.

Performance Evaluation of SVM based Abnormal Gait Analysis with Normalization

M. Pushpa Rani[1]
[1] Associate Professor in Computer Science,
Mother Teresa Women's University,
Kodaikanal, Tamil Nadu, India

G.Arumugam[2]
[2] Professor & Head, Dept. of Computer Science,
Madurai Kamaraj University,
Madurai, Tamil Nadu, India

***Abstract*—Support Vector Machine classifiers are powerful tools that are specifically designed to solve large-scale classification problems. In 1990s, Vapnik along with a group of other mathematicians and scientists developed a new statistical approach that is more efficient particularly in dealing with large classification problems which they called as Support Vector Machines (SVM). An SVM method is being broadly used in gait analysis because of its remarkable learning ability. In this paper, a two stage SVM algorithm is proposed for children abnormal gait analysis. The algorithm uses T-Test based preprocessing methods for feature selection, normalization and combines SVM for Classification. Only samples that have weak relationships with all the clusters are involved in SVM. Experimental results reveal that this algorithm based on T-Test-SVM combination achieves a remarkable recognition performance for children abnormal gait analysis with reduced Computational cost.**

Keywords Abnormal Gait Analysis, Support Vector Machine (SVM), Gait Data Classification, T-Test

I. INTRODUCTION

GAIT analysis is very significant for early diagnosis of gait diseases and treatment assessment. Doctors in earlier days used to diagnose gait diseases manually with the help of certain graphs generated by the gait analysis system, with which only vague clues which may or may not reflect the reality were obtained. In most cases, doctors had different views of opinion by interpreting these curves. As machine based learning technology has developed, it has gained much interest in gait analysis, which is of great support to doctors for more reliable and accurate diagnosis of a disease. A Gait analysis is a systematic study of human motion. i.e., walking, running, skipping and the like, which mainly concentrates on the physical activities. Gait analysis is very much useful to check out the human conditions; whether normal or abnormal using the eye and brain of the observers, augmented by instrumentation for measuring body movements, body mechanics and the activity of the muscles. In other words, gait analysis is used to assess, to plan and to treat the individuals with conditions affecting their ability to walk. There exist a lot of methods to check out these normality and abnormality on human gait, but SVM is found to be more suitable in terms of its efficiency [2]. Motion analysis provides large volume of gait data to describe the human motion; walking speed, joint angles, forces, and moments etc., Data like joint kinetics, joint moments and joint powers have also been used for gait recognition. Also this technology is very useful for checking the walking pattern of children as Children under the age of 13 have more chances to have different style of walking. To trace out any abnormality in children's walk, our proposed method takes a key role; and with this one could diagnose any existing fault features of walking in early ages of childhood itself. This will surely of great help for earlier treatment of gait abnormality in children.

An SVM method has been broadly used in gait analysis because of its remarkable learning ability, accuracy and efficiency. In this paper the SVM technique is performed by having a training set and test samples. The training set is categorized into different sets of conditions, which can be grouped into two classes i.e. normality and abnormality. For this, the collection of data is very important and several observations are needed. The information regarding leg length, height, cadence, stride length and age are some of the features which are of great help in this gait analysis. The SVM constructs a hyper plane or a set of hyper planes in a high or infinite dimensional space, which can be used for classification or clustering. In simple words, given a set of training examples, each with a label of belonging to one of these categories, an SVM training algorithm builds a model that predicts in which category a new example falls. The proposed method uses t-test-SVM for classification. Interestingly, a good separation is achieved by the hyper plane that has the largest distance to the nearest training data points of any class. The t-test method is used to normalize data prior to classification. In its simplest form t-test provides a statistical analysis of means of several groups and therefore can generalize the Student's two-sample *t*-test to more than two groups.

The following section of this paper is projected as follows: Section 2 discusses some of the related works done earlier in gait based classification. The proposed *t*-test SVM gait classification method is described in Section 3. Section 4 illustrates the performance Analysis and Section 5 concludes the paper with directions for future work.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

II. RELATED WORK

The study of human gait has created much interest in many application areas including biomechanics, clinical analysis, computer animation and biometrics. As a result many researches emerged in recent years and of which, a few studies related to Gait classification are noted herewith.

C. Bauckhage et al., [13] well thought-out about automatic gait analysis as a means to deduce if an observed walking pattern appears to be normal or not. When compared to most contributions to visual gait analysis, the problem dealt with the paper requires a representation that abstracts from individual gait characteristics but allows for the classification of gait across individuals. Addressing this requirement, the author presented a homeomorphism between 2D lattices and shapes that enables a robust vector space embedding of silhouettes. Sampling apt lattice points allows to roughly track the movement of limbs without requiring any limb recognition strategy. Combining shape representations obtained from several frames into lager feature vectors provides temporal context for the classification task. Experimental results expose a complete knowledge that gait classification using support vector machines yields excellent accuracy. Temporal filtering of the results of classification in further improvements of the reliability of the presented framework, because it lessens the effect of sporadic misclassifications.

A.H. Khandokerl et al., [3] demonstrated the effectiveness of wavelet based multi scale correlation exponents of MFC as features for automated screening an individual subject of proper balance control as being within healthy ranges, or having high enough risk to be categorized as a falls risk or a faller by using SVM. Findings of that study were based on a small sample of impaired subjects, compared to a relatively small sample of healthy peers. Therefore, further validation of the relative risk estimation task is suggested in a larger, more diverse sample of healthy and balance impaired falls risk in elderly adults, which may subsequently lead us to make more robust automated diagnostic model of falls risk estimation. The significance of this study is that it provides an early estimation of relative falls risk in the elderly that holds great potential for indicating balance improving interventions to reduce their relative risk of falls.

Jian Ni et al., [17] had a look about gait recognition, which is simulated in the small and medium-scale gait database. Higher recognition rate and faster recognition speed of the algorithm are verified. The reason that this algorithm obtains superior test results is: The paper adopts support vector machine based on hybrid kernel function. This method makes that the SVM model has better generalization ability. In the method of parameter selection, the text uses the objective function and combines OPS algorithm to select the best kernel parameter. The way combines the advantages of objective function and PSO algorithm to optimize SVM parameters. It significantly improves the optimization speed, at the same time obtains a good optimization effect.

J.H. Yoo et al., [18] had described an automated gait recognition system using back propagation neural network algorithm. Gait is the most common human motion, and each person appears to have his or her own characteristic gait pattern. To identify the human gait, a total of 27 parameters are considered as gait features. By calculating a class separability of the given feature, only 10 important features for classifying the gait are selected from these feature sets. Then, the enhanced back-propagation neural network algorithm is applied to the SOTON database, and recognition rate of 90% for 30 subjects is accomplished. The results achieved give promising performance and higher recognition rates than those of an earlier gait recognition approach.

Ju Han et al., [19] proposed a new spatio-temporal gait representation, called the Gait Energy Image (GEI), for individual recognition by gait. Different from other gait representations which consider gait as a sequence of templates (poses), GEI represents human motion sequence in a single image while preserving temporal information. To overcome the limitation of training templates, a simple model is proposed for simulating distortion in synthetic templates and a statistical gait feature fusion approach for human recognition by gait. Experimental results show that a) GEI is an effective and efficient gait representation and b) the proposed recognition approach achieves highly competitive performance with respect to the published major gait recognition approaches. This paper presents a methodical and comprehensive gait recognition approach, which can work just as fine as other complex published techniques in terms of effectiveness of performance while providing all the advantages associated with the computational efficiency for real-world applications.

Shakhnarovich et al. combined the face and MV-based gait. The front face was captured by one camera and the side-view of the person was captured by another camera. Face-alone, gait-alone and combined face and gait recognition rates were 80%, 87%, and 91%, respectively. Zhou et al. [21] used a single camera to capture both face and gait. Recognition rates for face and gait separately were 64.3% and 85.7%, a single respectively. Conversely, when they were combined, the recognition rate increased up to 100% [23]. In [22], WS-based gait recognition was combined with speaker verification. Performance proved to be appreciably better in a noisy environment, compared to when speaker verification was used alone. The EER was in the range of 2%-12%, less than half of the EER of individual modalities. In this group, gait is captured using a video-camera from distance. Video and image processing techniques are employed to extract gait features for recognition purposes.

BenAbdelkader et al. [25] used stride and cadence for person identification and verification. Johnson and Bobick [26] extracted static body parameters such as the height, the distance between head and pelvis, the maximum distance between pelvis and feet, and the distance between feet, and

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

used them for recognition. Most of the MV related gait recognition algorithms are based on the human silhouette [27, 28]. That is the image background is detached and the silhouette of the person is extracted and analyzed for recognition, For example, Liu and Sarkar [27] computed the average silhouettes over a gait cycle, and used the Euclidean distance between them to compute similarity.

III. METHODOLOGY

The block diagram of the proposed system is shown in Figure1. After suitable preprocessing, the Salient Gait Features are extracted from the possible Gait Signatures. These Gait features are then subjected to *t*-test normalization and subsequently to SVM Classifier.

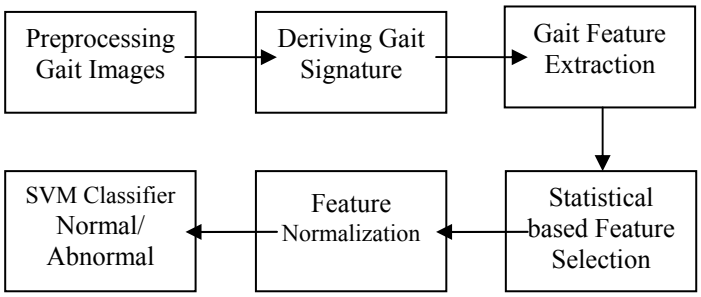


Figure1: Proposed Architecture

A. Statistical Methods for Gait Feature Selection & Normalization

The main objective of feature selection is to discover a subset of features, satisfying certain criteria. In pattern recognition, recognition metric will be the classification accuracy or inversely the classification error. But direct minimization of the classification error cannot be analytically performed, so a wide range of alternative statistics that are easier to evaluate are performed. The typical measure used in the gait feature selection is introduced as follows:

1. T-Test

The t-test finds whether the means of two groups are statistically dissimilar from each other. This analysis is appropriate to compare the means of two groups, and especially appropriate for the analysis of the two-group randomized experimental design.

The t-score (TS) [31] of feature i is defined as follows:

$$TS_i = max\left\{\left|\frac{\bar{x}_{ik} - \bar{x}_i}{m_k s_i}\right|, k = 1,2,\dots K\right\}$$

Where,

$$\bar{x}_{ik} = \sum_{j \in C_k} \bar{x}_{ij} / n_k,$$

$$\bar{x}_i = \sum_{j=1}^{n} x_{ij} / n_k,$$

$$s_i^2 = \frac{1}{n-K} \sum_{k} \sum_{j \in C_k} (x_{ij} - \bar{x}_{ik})^2,$$

$$m_k = \sqrt{\frac{1}{n_k} - \frac{1}{n}}$$

There are k classes.

$$max\{y_k, k = 1,2, \dots K\}$$

is the maximum of all k. C_k refers to class k that includes n_k, samples x_{ij} is the expression value of feature i in sample j x_{ik} is the mean expression value in class k for feature i. where n is the total number of samples. x_i is the general mean value for feature i. S_i is pooled within-class standard deviation for feature i. In fact, the TS used at this point is a t-statistic between the centroid of a specific class and the overall centroid of all the classes. Another possible model for TS could be a t-statistic between the centroid of a specific class and the centroid of all the other classes.

Two samples are given as input to the T-Test. The paired t-test determines whether input features differ from each other in a significant way under the assumptions that the paired differences are independent and identically normally distributed. This gives a clear view for Abnormal Gait Analysis

2. PCA (Principal Component Analysis)

Principal Component Analysis (PCA) involves a mathematical process that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. The fallouts of a PCA are usually discussed in terms of component scores and loadings.

PCA is the simplest of the proper eigenvector-based multivariate analyses. Often, its process can be thought of as revealing the internal structure of the data in a way which best explains the variance in the data. If a dataset (multivariate) is visualized as a set of coordinates in a high-dimensional data space (1 axis per variable), PCA supplies the user with a lower-dimensional picture, a "shadow" of this object when viewed from its (in some sense) most informative viewpoint. For a data matrix, X^T with zero empirical mean (the empirical mean of the distribution has been subtracted from the data set), where each row represents a different repetition of the

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

experiment, and each column gives the results from a particular probe, the PCA transformation is given by:

$$Y^T = X^T Y$$

$$= V\Sigma^T$$

where the m-by-n diagonal matrix Σ is an nonnegative real numbers on the diagonal and W Σ V^T is the Singular Value Decomposition (SVD) which is factorization of a rectangular real or complex matrix of X.

B Support Vector Machine

The support vector machine is used as a classifier in the paper. SVM is the one of the best linear classification method and kernel mixed applications. The SVM transforms the samples to high-dimension space by the kernel mapping, and then get the best linear classification surface of samples in this new space. This Non-linear transformation is achieved by appropriate inner product function. The best linear classification surface function of characteristics space can be described by the formula:

$$g(x) = \sum_{j=i}^{n} a_j y_i k(x, x_i) + b$$

Where (x_i, y_i) are the two types of sample collection divided in the sample space, b is the classification threshold, and k(x, x_i) is being the nonlinear kernel function that replaces characteristics space and meets Mercer conditions. Ascertain the best linear classification surface function is got by striking the best resolve a_i where i = 1, 2,…,n of the following function Q(a).

$$g(x) = \sum_{j=i}^{n} a_j y_i k$$

$$\max_a Q(a) = \sum_{i=0}^{n} a_i - 0.5 \sum_{i=0}^{n} \sum_{j=0}^{n} a_i a_j y_i y_j k(x_i, x_j)$$

$$\sum_{i=1}^{n} y_i a_i = 0, \qquad i = 1,2,...,n$$

$$0 \le a_i$$

The above equation is solving of quadratic function extreme value on condition that inequality, Q□(a) □□is convex function. Because its local optimal solution is global optimal solution, the solution is unique. Thus the best classification function of SVM is:

$$f(x) = \mathrm{sgn}(g(x)) = \mathrm{sgn}\left\{ \sum_{j=1}^{n} a_j y_j k(x, x_i) + b \right\}$$

$$\mathrm{sgn}\left\{ \sum_{i=1}^{n} a_i^* y_j^* k(x, x_i^*) + b \right\}$$

C SVM kernel functions

This work is the first attempt to test the classification ability of feature combinations in gait applications. We have used three main kernel functions for our study here. The Partial kernel function, which only influence to data near the test points and the kernel function more applied here is the Radial Basis Function:

$$k(x, x_k) = \exp\left\{ - |x - x_k|^2 / 2\delta^2 \right\},$$

where δ^2 is the width of the gaussian kernel. The overall kernel function which allows that, the data away from the test points will also have impact to kernel function. It is the polynomial kernel to be more suitable in this case:

$$k(x, x_{k)} = [(x, x_{k)} + 1]^{d,}$$

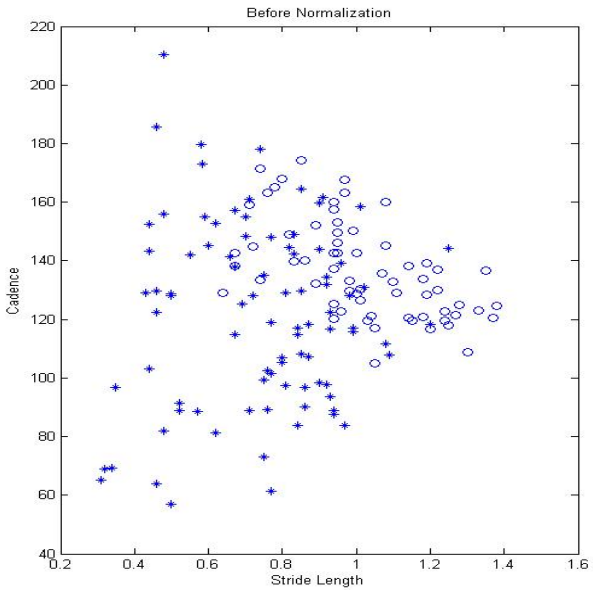
where d is degree of polynomial.

IV. Experimental Results

This study mainly deals with the performance analysis of the T-Test based SVM classification method for gait normality and abnormality. In this section, several experiments are carried out to test the validity of T-Test based SVM. A comparative analysis is also done for the proposed T-Test with PCA (Principal Component Analysis). The experimental data used in this study are obtained from the gait database of Virginia University [11]. There are totally 158 gait samples present in the database and all these samples are used for this experiment. These samples belong to 68 children with normal gait and 88 children with abnormal gait affected with cerebral palsy (CP). The ages of these children range from 2 years to 13 years. Four features of gait samples are selected for classification and they are stride length, cadence, leg length and age.

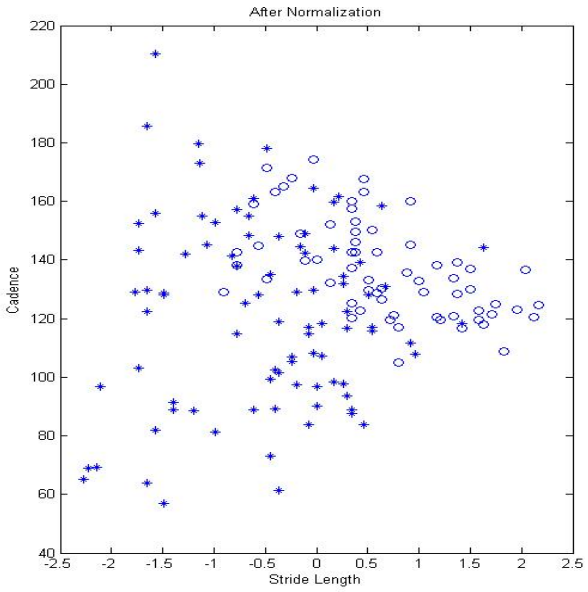
In this study, the t-test is applied to normalize the gait samples. Figure 2 shows the distribution of samples before and after normalization. As shown in Figure 2, the overlap of two sample sets is effectively reduced after normalization, which helps to improve the classification accuracy. Three kernel functions are used to build SVM classifiers in this study. By comparing the classification results of three classifiers, the most suitable kernel function may be decided for t-test-SVM. They are Radial Basis Function (RBF), linear and the polynomial. The RBF has best accuracy rate when compared to the other kernels such as the linear and the polynomial. In general, the RBF kernel is a reasonable first choice.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500



(a)

Figure 2: (a) Distribution of Raw Data Before Normalization.



(b)

(b) Distribution of Raw Data After Normalization.

This kernel nonlinearly maps samples of interest in a higher dimensional space so that, unlike the linear kernel, it can handle the case when the relation between class labels and attributes is nonlinear. Furthermore, the linear kernel is a special case of the RBF since the linear kernel with a penalty parameter C has the same performance as that of RBF kernel with some parameters. In addition, the sigmoid kernel behaves similar to the RBF for certain parameters. The next reason is the number of hyper parameters that influences the complexity of model selection. The polynomial kernel has additional hyper parameters than that of the RBF kernel.

Finally, the RBF kernel has fewer numerical difficulties. One key point should lie between 0 and 1, in contrast to polynomial kernels and linear kernel of which kernel values may go to infinity or zero, while the degree is large. Moreover, that the sigmoid kernel is not valid product of two vectors, under some parameters and conditions.

There are a few situations where the RBF kernel is not suitable. In meticulous, when the number of features is very large, one may just use the linear kernel.

By applying this kernel function, the accuracy of t-test -SVM and PCA SVM are compared at the end of this section.

Polynomial order *d* is an important parameter when polynomial kernel function is applied in t-test-SVM. The classification accuracy of gait samples by using different polynomial order is shown in Figure 3, and *d* is chosen from 1 to 10. As shown in Figure 3, the classification accuracy declines along with the increase of polynomial order. This is since the aspect of the feature space is high under a large polynomial order and it leads a declining generalization capability of SVM.

So, the linear kernel order *d* is set to be 1 in the following experiments. When *d* is 1, the polynomial kernel function actually is a linear kernel function with accuracy of 90.12%. But in case of the RBF kernel with $\sigma = 4$, $c = 100$,the absolute accuracy rate is of 98.15%, which leads the other two kernels.

When radial basis function (RBF) is applied, the kernel parameter σ in the RBF and the regularization parameter *C* may impact the classification accuracy of T-Test-SVM. Figure 4 elaborately shows the relationship between classification accuracy and parameters combination (d, C).

TABLE 1: BEST ACCURACY ACHIEVED FOR KERNEL FUNCTIONS

Kernel Function	Parameters	Accuracy (%)
Linear	d=1	90.12
Polynomial	d=1	85.69
RBF	σ=4, C=100	98.15

The classification accuracy of the RBF kernel function rates high of order 98%, this is shown in figure 4 for all the three cases of C=1, C=10, C=100.

The Figure reflects that, the generalization capability of the SVM enhances along with the increase of *C*. This is because the regularization parameter *C* may adjust the ratio of confidence interval and empirical risk.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

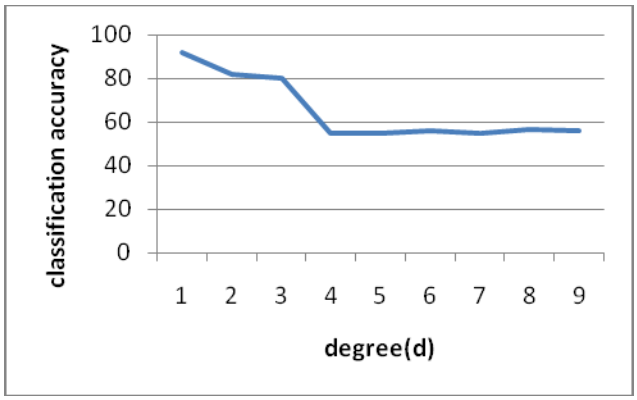


Figure 3, Classification Accuracy and Polynomial Order

TABLE 2: AVERAGE ACCURACY FOR VARIOUS σ & C VALUES

σ	C=1	C=10	C=100
0	20.12	54.21	92.07
1	35.08	63.75	94.29
2	54.82	78.64	94.38
3	65.54	85.11	95.93
4	75.12	90.39	96.17
5	90.43	92.42	97.59
6	92.54	92.78	97.26
7	93.72	96.64	98.07
8	92.05	96.89	98.15
9	92.34	96.92	98.15
10	92.16	96.11	98.05

In this case, the SVM has almost no change of the empirical risk and generalization capability. Table I shows the classification accuracy of the three classifiers.

TABLE 3: BEST ACCURACY ACHIEVED FOR VARIOUS METHODS

Algorithm	Accuracy (%)
Std SVM	87.68
PCA SVM	96.51
T-Test SVM	98.15

The generalization capability of the SVM is weak when *C* is small, because a small C indicates that the punishment for empirical risk is small and hence the possibility of the empirical risk is large. When C exceeds a certain value, the complexity of a classifier reaches the maximum allowed limit in the feature space.

Table 3 shows, that the T-Test SVM algorithm have better accuracy rate while compared with other algorithms. The accuracy rate of the PCA and T-Test SVM are compared with the different parameters for a particular dataset of children. These analysis reveals that T-Test have accuracy rate of 98.15%, while that of the PCA is comparably low.

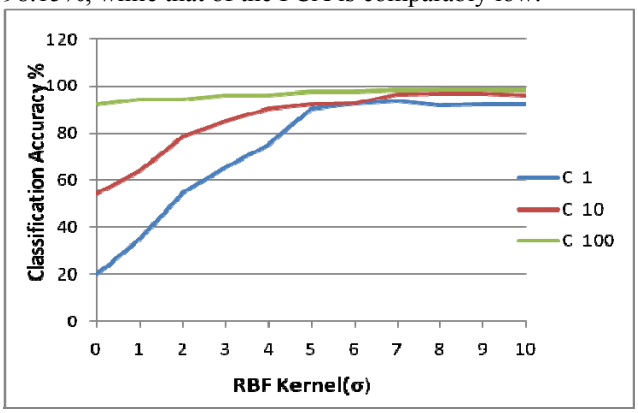


Figure 4, Classification Accuracy for Various Parameter Combinations

V. CONCLUSION

In this paper Abnormal Gait Analysis is done using the SVM with T-Test Combination. The results are compared with the other existing methods based on their Classification accuracy. An automated abnormal gait classification system is described using machine learning techniques. To achieve this, the gait signature has been extracted by combining a statistical approach and machine learning based analysis is further done using the anatomical knowledge. For the derived gait signatures, the motion parameters were calculated, and the gait features based on the motion parameters were extracted. The T-Test based SVM classifier is used to analyze the discriminatory ability of the extracted features. The result of the proposed method has produced very good classification rate which exceeds 98%. As such, the automated abnormal classification system not only accords with quantitative analysis in results, but also confirms distinctiveness as normal and abnormal gait. Hence this gait classification for medical diagnosis would be a real boon as its convenience will surely benefit the children and also the elderly. The drastic development of computer vision techniques also ensures that the clinical gait analysis put into practical realization may gradually be achieved.

REFERENCES

[1] Mostayed, A. Mynuddin, M. Mazumder, G. Sikyung Kim Se Jin Park," Abnormal Gait Detection Using Discrete Fourier Transform", International Conference on Multimedia and Ubiquitous Engineering, 2008.

[2] Bauckhage, C. Tsotsos, J.K. Bunn, F.E.,"Detecting Abnormal Gait", Proceedings of Canadian Conference on Computer and Robot Vision, 2005.

[3] A.H. Khandokerl, Daniel Lai, Rezaul K. Begg and Marimuthu Palaniswami,"A Wavelet Based Approach for Screening Falls Risk in the Elderly using Support Vector Machines", Fourth International Conference on Intelligent Sensing and Information Processing, 2006. ICISIP 2006.

[4] Inman, V. T., Ralston, H. J., and Todd, F.: Human Walking. Williams & Wilkins, Baltimore (1981).

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[5] Murray, M. P., Drought, A. B., and Kory, R. C.: Walking Patterns of Normal Men. Journal of Bone and Joint Surgery, 46A(2) (1964) 335-360.

[6] Johansson, G.: Visual Perception of Biological Motion and a Model for Its Analysis. Perception and Psychophysics, 14(2) (1973) 201-211.

[7] Winter, D. A.: The Biomechanics and Motor Control of Human Gait: Normal, Elderly and Pathological. Waterloo Biomechanics, Ontario (1991).

[8] Kyoungchul Kong Tomizuka, M," A Gait Monitoring System Based on Air Pressure Sensors Embedded in a Shoe", IEEE/ASME Transactions on Mechatronics, 2009.

[9] Nixon, M. S., Cater, J. N., Grant, M. G., Gordon, L., and Hayfron-Acquah, J. B.: Automatic Recognition by Gait: Progress and Prospects. Sensor Review, 23(4) (2003) 323-331.

[10] J. Devore and R. Peck, Statistics: The Exploration and Analysis of Data, third ed. Duxbury Press, 1997.

[11] IEEE transactions on rehabilitation engineering, vol. 5, no. 4, december 1997

[12] Foster, J. P., Nixon, M. S., and Prügel-Bennett, A.: Automatic Gait Recognition using Area-based Metrics. Pattern Recognition Letters, 24(14) (2003) 2489-2497.

[13] Christian Bauckhage, John K. Tsotsos, Frank E. Bunn, "Automatic Detection of Abnormal Gait", International journal of Image and Vision Computing, 2007.

[14] Zongyi Liu and Sudeep Sarkar, "Improved Gait Recognition by Gait Dynamics Normalization," IEEE Trans on Pattern Analysis and Machine Intelligence, Vol. 28, No. 6, June 2006.

[15] L. Lee, G. Dalley, and K. Tieu, "Learning Pedestrian Models for Silhouette Refinement," Proc. Int'l Conf. Computer Vision, pp. 663-670, 2003.

[16] A. Veeraraghavan, A.R. Chowdhury, and R. Chellappa, "Matching Shape Sequences in Video with Applications in Human Movement Analysis," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 12, pp. 1896-1909, Dec. 2005.

[17] Jian Ni, Libo Liang, " Gait Recognition Method Based on Hybrid Kernel and Optimized Parameter SVM", IEEE International Conference on Computer Science and Information Technology, 2009. ICCSIT 2009.

[18] Jang-Hee Yoo, Doosung Hwang," Automated Human Recognition by Gait using Neural Network", IEEE Conference on Image Processing Theory, Tools & Applications, 2008.

[19] Ju Han, Bir Bhanu, "Individual Recognition Using Gait Energy Image", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, No. 2, February 2006.

[20] G. Shakhnarovich, L. Lee, and T. Darrell. Integrated face and gait recognition from multiple views. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition., 2001.

[21] Xiaoli Zhou, Bir Bhanu, and Ju Han. Human recognition at a distance in video by integrating face profile and gait. In 5th International Conference on Audio- and Video-Based Biometric Person Authentication, pages 533–543, July 2005.

[22] Elena Vildjiounaite, Satu-Marja Makela, Mikko Lindholm, Reima Riihimaki, Vesa Kyllonen, Jani Mantyjarvi, and Heikki Ailisto. Unobtrusive multimodal biometrics for ensuring privacy and information security with personal devices. In Pervasive, pages 187–201, May 2006. Springer LNCS.

[23] G. Shakhnarovich, L. Lee, and T. Darrell. Integrated face and gait recognition from multiple views. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2001.

[24] Davrondzhon Gafurov, "A Survey of Biometric Gait Recognition: Approaches, Security and Challenges", NIK-2007 conference.

[25] C. BenAbdelkader, R. Cutler, and L. Davis. Stride and cadence as a biometric in automatic person identification and verification. In Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pages 357–362, May 2002.

[26] Amos Y. Johnson and Aaron F. Bobick. A multi-view method for gait recognition using static body parameters. In Third International Conference on Audio- and Video-Based Biometric Person Authentication, pages 301–311, June 2001.

[27] Zongyi Liu and Sudeep Sarkar. Simplest representation yet for gait recognition: Averaged silhouette. In International Conference on Pattern Recognition, pages 211–214, 2004.

[28] Zongyi Liu, Laura Malave, and Sudeep Sarkar. Studies on silhouette quality and gait recognition. In Computer Vision and Pattern Recognition, pages 704–711, 2004.

[29] Yanmei Chai, Jinchang Ren, Rongchun Zhao, and Jingping Jia. Automatic gait recognition using dynamic *va*riance features. In International Conference on Automatic Face and Gesture Recognition, pages 475 – 480, 2006.

[30] http://www.wrongdiagnosis.com/symptoms/abnormal_gait_in_children/common.htm.

[31] M. Pushpa Rani and G.Arumugam, "An Efficient Gait Recognition System For Human Identification Using Modified ICA ",International Journal of Computer Science & Information Technology, vol. 2, No 1, pp 55-67,January 2010.

AUTHORS PROFILE

Ms.M.Pushpa Rani is an Associate Professor in Computer Science at Mother Teresa Women's University, Tamil Nadu, India. She received her Master's degree in Computer Applications from Bharathiar University, Coimbatore, India, and is currently pursuing doctoral Research in Madurai Kamaraj University, Madurai, India. She has published many articles in International Journals and more than 30 papers in National and International Conferences. Her current research areas include Image Processing, Biometrics and Adaptive Learning System.

Dr G Arumugam is a senior Professor in the Department of Computer Science, Madurai Kamaraj University, Madurai, He did his Masters degree in Applied Mathematics specializing in Computer Science in the PSG College of Technology, Coimbatore in the year 1980. He started his research carrier from the Department of Mathematics, IIT Kanpur in 1981.He got his Doctorate degree from the University of Pierre and Marie Curie , France in 1987.He was a Post-Doctoral Fellow in the University of JYVASKYLA, Finland for a period of three months.

He worked in the Hexaware Info Systems, Chennai as a Project Manager and as a Consultant in the Polaris Software, Chennai to gain industrial experience.. He was in Singapore for a period of 2 ½ years as Visiting Professor in the Ngee Ann Polytechnic, Singapore from July 2000 to Nov.2002.He has published several papers in the international and national journals and in the conference proceedings. His area of research is design and analysis of algorithms, data mining, and cryptography and network security.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Design of a Dynamic Intelligent Intrusion Detection System Model

Vivian Ogochukwu Nwaocha
School of Science and Technology
National Open University of Nigeria, Lagos
webdevee@yahoo.com

Abstract*—Most industries require the internet in order to function effectively. We do our banking online, purchase items online and store a wealth of personal information on the hard drive. Unfortunately, thieves know this, and go to great lengths to get our information. As a growing number of companies are connected over the Internet to branch offices in other parts of the country or world, they expose themselves to potential attacks from other individuals or companies after their trade secrets. This could result in losses in the millions or even billions, if important, confidential information is stolen and ends up in the wrong hands. Thus, several organizations invest a great deal in network security in order to ensure that their systems do not fall prey to attacks. However, the existing intrusion detection systems are restricted in their monitoring functionality and require frequent updates and patches. This study presents a dynamic Intelligent Intrusion Detection System model. The proposed model applies fuzzy logic and data mining techniques and a simultaneous implementation of counter measures as the effective means of dynamic intrusion detection.*

Keywords-arificial intelligence; business, dynamic intelligent intrusion detection system model; information; network; organization; security;

I. INTRODUCTION

The widespread use of information stored and processed on network-based systems in most businesses has increased the necessity of protecting these systems. Most businesses are constantly experiencing new threats and vulnerabilities in their applications. Therefore, trying to keep up with emerging threats, applying patches against known vulnerabilities, updating antivirus software, updating firewall rules and all of the other security measures can have a network or security administrator working 18 hour days, 7 days a week with no vacation (Bradley, 2004). There is a crucial need to address security issues that affect networks. It is equally important to be able to sift through the mountains of potential threats and determine which ones truly affect your network so that your time and resources can be put to the most efficient use.

An Intrusion Detection System (IDS) is a device or software application that monitors network and/or system activities for malicious activities or policy violations and produces reports to a Management Station (Scarfone et al. 2007). Intrusion Detection Systems (IDS) are defined by both the method used to detect attacks and the placement of the IDS on the network. IDS may perform either misuse detection or anomaly detection and may be deployed as either a network-based system or a host-based system. These results in four general groups: misuse-host, misuse-network, anomaly-host and anomaly-network. Typically, Intrusion Detection System, can detect, prevent and react to the attacks. Intrusion Detection has become an integral part of the information security process. However, it is not technically feasible to build a system with no vulnerabilities; as such, intrusion detection continues to be an important area of research. Diverse techniques to Intrusion Detection are currently being used, but they are comparatively unproductive. Artificial Intelligence (AI) plays a driving role in protection services. In this study, a dynamic Intelligent Intrusion Detection System model, based on specific AI approach for intrusion detection is employed. The method that is being explored comprise fuzzy logic with network profiling, which applies simple data mining techniques to process the network data. The proposed hybrid system combines anomaly and misuse detection. By applying this technique, suspicious intrusions can be traced back to its original source and any traffic from that particular source will be redirected back to them in future.

II. OBJECTIVE

This study seeks to design and develop a dynamic intelligent intrusion detection system that would be accurate, low in false alarms, not easily cheated by small variations in patterns, adaptive and respond in real-time.

III. LITERTURE REVIEW

Intrusion Detection Systems have in fact been around for a while and a number of studies have been carried out in this area to enhance its efficacy. Basically, they come in a variety of "flavors" and approach the goal of detecting suspicious traffic in different ways. They are characterized by the method employed to detect attacks and the placement of the Intrusion Detection System (IDS) on the network. Typically, Intrusion Detection Systems (IDSs) may execute either misuse detection or anomaly detection and may be deployed as either a network-based system or a host-based system. Thus, they can be categorized into four major classes: misuse-host, misuse-network, anomaly-host and anomaly-network.

Misuse detection relies on matching known patterns of hostile activity against databases of past attacks. They are highly effective at identifying known attack and vulnerabilities, but rather poor in identifying new security threats. Anomaly detection will search for unusual occurrences by applying statistical measures or artificial intelligence methods to compare current activity against historical knowledge. Common problems with anomaly-based systems

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

are that, they often require extensive training data for artificial learning algorithms, and they tend to be computationally expensive, because several metrics are often maintained, and need to be updated against every systems activity. In 1998, Lee, Stolfo S. and Mok K, carried out a research on mining audit data to build intrusion detection models. Mukkamala, Gagnon and Jaiodia explored the integration of data mining techniques with intrusion detection methods (Mukkamala et al. 2000). In 2001, Stolfo S., Lee and Chan studied "Data mining-based Intrusion detectors" in order to produce a summary of the Columbia Intrusion Detection System. Data mining techniques have equally been employed in mining normal patterns for audit data. In his work, Lunt (1993) designed an Intrusion Detection Expert System that encodes an expert's knowledge of known patterns of attack and system vulnerabilities as if-then rules.

Some techniques applying artificial neural networks in the intrusion detection system have been proposed. Debar, Becker and Siboni, designed a neural network component for an intrusion detection system (Debar et al, 1992). Tan carried out a research on the application of neural networks to UNIX Computer security (Tan, 1995). In 2004, Wang J., Wang Z. and Dai, designed a network intrusion detection system based on artificial neural networks. NeGPAIM (Botha et al., 2002) applies trend analysis, fuzzy logic and neural networks to minimize and control intrusions. Intrusion detection systems that are currently available, particularly commercial intrusion detection systems that must resist intrusion attacks are based on misuse detection approach, which implies that these systems will only be able to spot known attack types and in most cases they tend to be ineffective due to various reasons like non-availability of attack patterns, time consumption for developing new attack patterns, insufficient attack data etc.

Compared to the classical approach, several benefits arise from applying fuzzy methods for the development of Intrusion Detection System (IDS). Consequently, Fuzzy logic systems have been employed in the computer security field since the early 90's. This technique provides some flexibility to the uncertain problem of intrusion detection and allows much greater complexity for IDS. Nevertheless, most of the fuzzy IDS require human experts to determine the fuzzy sets and set of fuzzy rules. These tasks are time consuming. However, if the fuzzy rules are automatically generated, less time would be consumed for building a good intrusion classifier and shortens the development time of building or updating an intrusion classifier. In their studies, Dokas, Ertoz, Vipin, Srivatava and Tan (Dokas et al., 2002) suggested a model for building rare class prediction models for identifying known intrusions and their variations and anomaly/outlier detection schemes for detecting novel attacks whose nature is unknown.

The most recent in fuzzy is to use the Markov model. As suggested in Xi et al (2004), a Window Markov model is proposed, the next state in the window equal evaluation to be the next state of time t, so they create Fuzzy window Markov model. As discussed in the work of Gomez and Dasgupta (2002), a technique to generate fuzzy classifiers using genetic algorithms that can detect anomalies and some specific intrusions was proposed. The main idea is to evolve two rules, one for the normal class and other for the abnormal class using a profile data set with information related to the computer network during the normal behavior and during intrusive (abnormal) behavior. Fuzzy preference relation is another method applied to intrusion detection based on fuzzy satisfaction function. This is applied for comparison of attack signatures. Fuzzy signatures (their gamma resolution sets) are combined by fuzzy operators (Manic and Wilamowski, 2001). Yao, Zhao and Saxton in their studies (Yao et al., 2003) proposed a dynamic approach that tries to discover known or unknown intrusion patterns which uses Support Vector Machine (Chen and Wang,2003).A dynamic fuzzy boundary is developed from labeled data for different levels of security needs.

IV. THE PROPOSED FRAMEWORK

This study presents an Intelligent Intrusion Detection System model which utilizes fuzzy logic along with data mining technique. The proposed model is the modified version of FIRE (Dickerson & Dickerson) system. In order to discover the features that would facilitate the detection of attacks the FIRE system employs a simple data-mining algorithm. The security administrator uses the fuzzy sets produced by the system to generate fuzzy rules. Conversely, a mechanism to automate the rule generation process and reduce the human intervention is proposed. Artificial Intelligence techniques have also been explored to build intrusion detection systems based on knowledge of past behavior and normal use. They have shown potentiality for anomaly detection with limited ability.

In this model, SNORT (Snort, 2010), an open source network intrusion prevention and detection system (IDS/IPS) developed by Sourcefire is equally employed. It is a famous open source packet sniffer. The data processor and classifier summarizes and tabulates the data into carefully selected categories i.e. the attack types are carefully correlated. This is the stage where a kind of data mining is performed on the collected data. In the next stage, the current data is compared with the historical mined data to create values that reflect how new data differs from the past observed data. The inference engine works based on Mandami inference mechanism. Based on outcome of previous studies, it can be concluded that this mechanism would suit the research requirements.

Based on the facts from the analyzer, the decision will be taken whether to activate the detection phase or not. If the detection phase is activated then an alert will be issued and the tracer phase will be initiated. This phase will trace back to the intruders original source address location. A framework for tracing the abnormal packets back to its original source based on single packet is proposed. This tends to be the most tedious phase of the project. Once the original path has been identified and verified then all the attacks from that particular host will be redirected to their source in future.

SNORT_INLINE (Snort, 2010) has been proven to be the best in changing the appropriate packet values. Figure 1 below presents the architecture of the proposed model.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

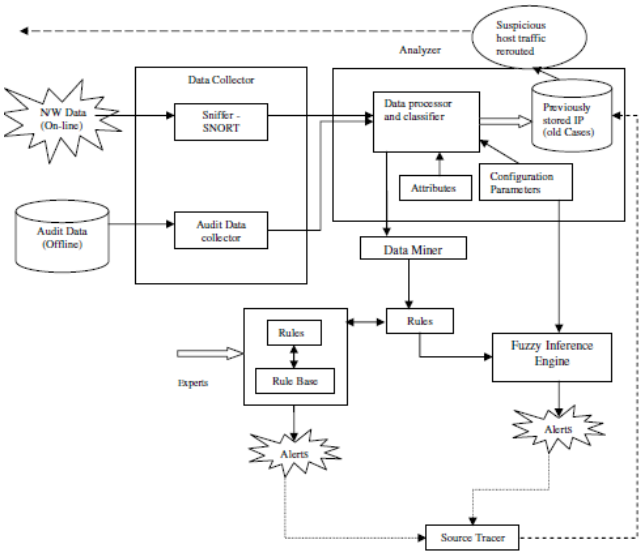


Figure 1: The Framework of a Dynamic Intelligent Detection System Model

The proposed model is a hybrid which comprises fuzzy logic with data mining to provide a more efficient anomaly and misuse intrusion detection in a real-time environment. Attributes representing relevant features of the input data have to be established prior to any data analysis. Once relevant attributes have been defined and data source identified, the Data Analyzer computes configuration parameters that control the operation of the IDS. This module analyzes packets and computes aggregate information by grouping packets. These packets can be placed in permanent size groups (s−group) or in groups of packets captured in a fixed amount of time (t−group). Each s−group contains the same number of packets covering a variable time range and each t−group contains a variable number of packets captured over a fixed period of time. Rules are expressed as a logic implication p->q, where p is called the antecedent of the rule and q is called the consequence of the rule.

In order to implement the Data Miner a variation of Kuok's algorithm is employed (Kuok et al., 2002). This approach allows for efficient, single–pass, record processing by partitioning data into hierarchical files. To facilitate the discovering of association rules for binary, categorical and numerical attributes, the Apriori is incorporated to Kuok's algorithm. The final output of the algorithm is a set of rules that meet the confidence and support constraints given as input. The tracing of the source of the packet is needed for obtaining the exact information about the intruder. The IDS will be providing information that an exceptional event has occurred, the packet and the time of the attack. Once a trace back is requested, a query message consisting of the packet, egress point and the time of receipt is sent to all the Local Data Managers (LDM).Time is critical as this must take place while the appropriate values are still resident at the DC (Data Collector). Once the values are safely transferred to TM, the trace back process will no longer be under real-time constraints. Local Data Manager is responsible for a particular network. Later LDM responds with the partial attack graph and the packet as it entered the region. The attack graph either terminates within the region managed by the LDM, in which a source has been identified, or it contains nodes to the edges of the other LDM network region. Next, TM sends a query to the LDM adjacent of that edge node.

V. CONCLUSION AND FUTURE WORK

The goal of this paper is to contribute to the development of network security techniques in general and Network Intrusion Detection System (NIDS) research specifically. The outcome of this study would thus facilitate the task of Network Administrators in particular as well as ensure that businesses and organisations secure their information in order to perform their operations effectively. In this preliminary study a dynamic Intelligent Intrusion Detection System Model has been designed and developed. Nevertheless, the implementation of the overall model has not been done. The future work will thus entail deploying this hybrid system model in a real-time environment as well as applying clustering and classification algorithm in the data mining process to obtain top quality results.

REFERENCES

[1] Botha M., Solms R., Perry K., Loubser E., Yamoyany G., "The utilization of Artificial Intelligence in a Hybrid Intrusion Detection System", SAICSIT, 149-155 2002

[2] Bradley T., "Evaluating Threats to your Network". Former Internet/Network Security Guide, 2004

[3] Debar, Becker M., Siboni D., "A neural network component for an intrusion detection system." IEEE Computer Society Symposium on Research in Computer Security and Privacy, 240-250 1992

[4] Dickerson J E., Dickerson J A., "Fuzzy Network Profiling for Intrusion Detection" Proceedings of NAFIPS 19th International Conference of the North American Fuzzy Information Processing Society, Atlanta 2000.

[5] Dokas P., Ertoz L., Vipin Kumar., Srivastava J., Tan P., "Data Mining for Network Intrusion Detection". National Science Foundation Workshop on Next Generation Data Mining, USA 2002

[6] Gomez J., Dasgupta D., "Evolving Fuzzy classifiers for Intrusion Detection". Proceedings of 2002 IEEE Workshop in Information Assurance, USA NY 2002

[7] Kuok C., Fu A., Wong M., "Mining fuzzy associationrules in databases" SIGMOD Record 17 (1) 41-46, 2002.

[8] Lunt. T. "Detecting intruders in computer systems". Conference on auditing and computer technology, 1993

[9] Lee, Stolfo S., Mok K., "Mining audit data to build data to build intrusion detection models." Fourth international conference on knowledge discovery and data mining, New York, AAAI Press 66-72, 1998

[10] Manic M., and Wilamowski, "Fuzzy preference approach for Computer Network Attack Detection". IEEE conference on Fuzzy systems 2001

[11] Mukkamala, R., Gagnon J., Jaiodia S., "Integrating data mining techniques with intrusion detection methods". Research Advances in Database and Information systems security, 33-46, 2000

[12] Scarfone, Karen; Mell, Peter (February 2007). "Guide to Intrusion Detection and Prevention Systems (IDPS)". Computer Security Resource Center (National Institute of Standards and Technology) (800-94).

[13] SNORT www.snort.org , 2010

[14] SNORT INLINE http://snort-inline.sourceforge.net/

[15] Stolfo S., Lee, Chan. "Data mining-based Intrusion detectors : An overview of the Columbia IDS" Project SIGMOD Record Vol 30, No 4, 2001

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[16] Tan K., "The Application of Neural Networks to UNIX Computer security". IEEE International conference on Neural Networks Vol 1, 476-481 1995

[17] Xi Z., Sun J., Wenjie L., "Intrusion Detection using Fuzzy Window Markov Model". CCECE Niagra Falls, Canada 2004

[18] Yao TJ, Zhao, Saxton A study on fuzzy intrusion detection, Unpublished work, University of Regina, Canada, 2003

[19] Y. Chen and J.Z. Wang, "Support Vector Learning for Fuzzy Rule-Based Classification Systems", IEEE Transaction on Fuzzy Systems, 2003

AUTHORS PROFILE

Vivian Ogochukwu Nwaocha is currently involved in coordinating Computer Science and Information Technology programs at the National Open University of Nigeria. Her main research interests are network security, artificial intelligence, mobile learning and assistive technologies. A good number of research papers authored by Vivian have been published in various local and international journals. Vivian has equally written a number of books which are accessible online. She has participated in several community and service development projects in Nigeria and beyond. Vivian is a member of Computer Professionals Registration Council of Nigeria, Nigeria Computer Society, Prolearn Academy, elearning Europe, IAENG society of Computer Science, IAENG society of Artificial Intelligence, IAENG society of Bioinformatics and several online social networking communities.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Multipath Routing for Self-Organizing Hierarchical Mobile Ad-hoc Networks – A Review

Udayachandran Ramasamy
Professor, Department of Computer Science and Engineering
Sri Ramakrishna Institute of Technology
Coimbatore, Tamilnadu, India
ruday2004@yahoo.co.in

Sankaranarayanan K.
Principal
Akshaya College of Engineering and Technology
Coimbatore, Tamilnadu, India
kkd_sankar@yahoo.com

***Abstract*— Security has become a primary concern for providing protected communication between mobile nodes in a hostile environment. The characteristics of Ad-hoc networks (dynamic topology, infrastructure less, variable capacity links, etc) are origin of many issues. Limited bandwidth, energy constraints, high cost security are the encountered problems. This type of networks pose particular challenges in terms of Quality of Service (QoS) and performance. In this paper, the issues of multipath routing in MANETs are reviewed and performances of such MANETs are compared to discuss the application of multipath routing and its effects on different layers to support QoS.**

Keywords- MANET, Mobile Ad-hoc networks and QoS.

I. *INTRODUCTION*

In Adhoc networks, more flavours of routing options are provided to cover the flexible nature of network elements. This paper is organized as follows: In section II different categories of MANET's Routing protocols are classified. In Section III multipath routing will be discussed in depth and those Adhoc protocols offering multipath routing will be addressed in detail. Conclusion and References will be mentioned in Section IV and V.

II. MOBILE AD-HOC NETWORKS AND PROTOCOLS

The major challenges that a routing protocol designed for adhoc wireless networks faces are mobility of nodes, resource constraints, error-prone channel state, and hidden and exposed terminal problems[1][2]. In most MANETs, multipath protocols are needed to facilitate efficient connectivity between senders that are not necessarily within each other's wireless range. MANET routing protocols are divided into the following categories.

- Flat Routing Protocols
  - Proactive Routing (Table-Driven)
  - Reactive Routing (On-Demand)
  - Hybrid Routing (blend of Reactive and Proactive)
- Hierarchical (Zonal/Cluster-Based) Routing Protocols.
- Geographic Position Assisted Routing Protocols
- Power-Aware Routing Protocol
- Security- Aware Routing Protocol
- Multicasting Routing Protocols
  - Geographical Multicast (Geocasting)
  - Tree-based
  - Mesh-based
  - Zone Routing
  - Associativity-Based
  - Differential- Destination
  - Weight-Based
  - Preferred Link-based

The goals of these protocols could be summarised as [3]:

- Minimal Control Overhead
- Minimal Processing Overhead
- Multihop Routing Capability
- Dynamic Topology Maintenance
- Loop Prevention

III. MULTIPATH ROUTING PROTOCOLS

There exists a set of QoS routing and best-effort routing protocols that use multiple paths between a source-destination pair. There are several advantages in using multipath routing. Some of the advantages [1] include the reduction in route computing time, the high resilience to path breaks, high call acceptance ratio, and better security. For TCP, these advantages may add to throughput degradation. These can lead to significant amount of out-of-order packets, which in turn generates a set of duplicate acknowledgements (DUPACKS) which cause additional power consumption and invocation of congestion control.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A. Multipath routing in Reactive Protocols

MSR 'Multipath Source Routing Protocol' [4] that supports multipath routing is a direct descendant of DSR. In MSR, the throughput, end-to-end delay, and drop rate have been improved greatly. Another routing protocol AOMDV 'On-Demand Multipath Distance Vector Protocol' [5], extends the single path AODV protocol to compute multiple paths. Under wide range of mobility traffic scenarios, AOMDV offers a significant reduction in delay and up to 20% reduction in routing load and the frequency of route discoveries.

B. Multipath Routing in Proactive Protocols

DSDV 'Destination-Sequenced Distance- Vector Routing' [6], maintain routing updates among all nodes all the time. In fact, many proactive protocols tend to offer shortest path to each destinations. This is done by continuously monitoring the network topology. Unlike reactive routing algorithms, proactive routing protocols are capable of repairing broken routes in a short time. This is done by collecting network topology continuously. The drawback of DSDV however is the requirement of parameters such as the periodic update interval, maximum value of the 'settling time' for a destination and the name of update intervals, which may become known before a route is considered stale. TERA 'Tree Exchange Routing Algorithm' is an extension to standard distance vector routing algorithms, which is based on multipath.

C. Multipath Routing in Hybrid Protocols

'ZRP' Zone Routing Protocol is a hybrid routing protocol incorporate the merits of both on-demand and proactive routing protocols. ZRP is similar to a cluster with the exception that each node acts as a cluster head and a member of other clusters. The routing zone forms few mobile adhoc nodes within one, two or more hops away where the central node is located.

Ant Hoc Net [7] 'Ant Agents for Hybrid Multipath Routing in Mobile Adhoc Networks' the route setup of this scheme is performed by reactive algorithm and the route probing and exploration are done by proactive scheme. This scheme can outperform AODV in terms of delivery ratio and average delay, especially in more mobile and large networks. Scalability is also promising in this scheme. However, relatively large amount of overhead could be mentioned as a drawback and also less adaptability to the network situation.

D. Multipath Routing in Hierarchical Protocols

The use of routing hierarchy has several advantages, the most important ones being reduction in the size of routing tables and better scalability. The Hierarchical State Routing (HSR) protocol [8] is a distributed multilevel hierarchical routing protocol that employs clustering at different levels with efficient membership management at every level of clustering. The use of clustering enhances resource allocation and management.

Though the reduction in the amount of routing information stored at nodes is appreciable, the overhead involved in exchanging packets containing information about the multiple levels of hierarchy and the leader election process makes the protocol unaffordable in the adhoc wireless networks context.

The Fisheye State Routing Protocol (FSR) [8] is a generalization of GSR [9] protocol. FSR uses the fisheye technique to reduce information required to represent graphical data, to reduce routing overhead. The basic principle behind this technique is the property of a fish's eye that can capture pixel information with greater accuracy near its eye's focal point. This accuracy decreases with an increase in the distance from the centre of the focal point. This property is translated to routing in adhoc wireless networks by a node, keeping accurate information about nodes in its local topology, and not-so-accurate information about far-away nodes, the accuracy of the network information decreasing with increasing distance. Hierarchical Max-Flow Routing 'HMFR' forwards packets in such a way that the impact of failures is minimized. However, the computational complexity of Max-flow routing is quite high, making it not acceptable for moderate size networks.

E. Multipath Routing in Geographic Position Assisted Routing Protocols

Multipath location - Aided Routing 'MLAR' is a multipath version of LAR [10]; that uses position information (2D or 3D) to make routing decisions at each node. The proposed algorithm in [11] uses a 3D approach which is a new hierarchical, Zone based 3D routing algorithm based on GRID [12]. The approach proposes a replacement of LAR with Multipath LAR (MLAR) in GRID. It is expected to have significant performance differences in 3D and as to whether single or multipath algorithms should be used in a particular scenario. The simulation results demonstrate the performance benefits of MLAR over LAR and AODV in most mobility situations. AOMDV delivers more packets compared to MLAR,but at a cost of more frequent flooding to control packets and thus bandwidth usage than MLAR.

F. Multipath Routing in Power-Aware Protocols

The limitation on the availability of power for operation is a significant bottleneck, given the requirements of portability, weight, and size of commercial hand-held devices. Hence the use of routing metrics that consider the capabilities of the power sources of the network nodes contributes to the efficient utilization of energy and increase the lifetime of the network. Singh et al [13] proposed a set of routing metrics that supports conservation of battery power. The routing protocols that select paths so as to conserve power must be aware of the states of the batteries at the given node as well as at the other intermediate nodes in the path.

An interesting insight of power-aware adhoc protocols has been presented in [14] in which optimization at the network layer is of major concern. The research addresses power control, routing, and sleep mode (stand-by) control.

Multipath Power Sensitive Routing Protocol 'MPSR' [15] is another adhoc routing protocol with interest in power aware communication. MPSR shows how an efficient heuristic-based multipath technique can improve the mean-time-to-node failure and maintain the variance in power of all the nodes as low as possible. The simulation results show performance optimized in MPSR protocol compared to the Dynamic Source Routing 'DSR'.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

G. Multipath Routing in Multicasting Protocols

Based on the type of operation, multicast protocols for adhoc networks are broadly classified into two types: Source-initiated protocols and receiver-initiated protocols. There exists certain other multicast protocols such as MCEDAR [16] and AM Route [17] which may not strictly fall under the above two types. Multicast Routing Protocols are of great interest as the demand for such communication is on the rise. Multipath Multicast Routing Algorithm 'MRPM' quick distributed dynamic algorithm can manage network resources efficiently.

Multi-Flow Real-Time Transport protocol 'MRTP' [18] is another example of a mesh-based adhoc-based protocol that offers multipath routing for multicast application. MRTP is motivated by the observations of effective path diversity in combating transmission errors in adhoc networks and effective data partitioning techniques in improving the queueing performance of real-time traffic. The simulation results show performance improvement in lost packets per frame and better management.

Multi-Objective Multipath Routing Algorithm for Multicast Flows 'MMRAM' [19] proposes a multi-objective traffic engineering scheme using different distribution trees to multicast several flows. MMRAM is an attractive candidate for Multi Protocol Label Switching 'MPLS'. This multi-tree routing protocol uses a multicast transmission with load balancing.

H. Multipath routing in Security Protocols

The existence of multiple paths between nodes in an Adhoc network is exploited to increase the robustness of the data confidentiality [20]. The proposed algorithm is tested against time for intrusion defection and robustness.

TABLE I. ADHOC ROUTING PROTOCOLS DISCUSSED

ROTOCOLS	
DSR	Reactive
MSR	Reactive
AODMDV	Reactive
AODV	Reactive
DSDV	Proactive
TERA	Proactive
HMFR	Hierar
ZRP	Hybrid
ANT.NET	Hybrid
HSR	Hierar
ZHLS	Hierar
CGSR	Hierar
MLAR	Geog
MPSR	Power
MRPM	Multicast
MRTP	Multicast
MMRAM	Multicast
MTS	Security

Another multipath routing algorithm for data security enhancement, Multipath TCP security 'MTS', is discussed in [21]. In MTS, the source node chooses the available routes adaptivity rather than testing the 'Stored routes' one by one exhaustively. Compared to AODV and DSR, MTS has a better number of participating nodes and highest interception ratio. The average end-to-end delay between MTS, AODV and DSR shows that beyond speeds of 1.7m/s, MTS delay drops rapidly and performs better in respect to the other two routing protocols. The protocols discussed are tabulated in Table 1.

CONCLUSIONS

Multipath routing was the main focus of this paper and we investigated its effects of multipath routing in variety of protocols. In all these, performance enhancements were observed and promising results pointed to the better deployment of the schemes when multipath routing is used.

REFERENCES

[1] Siva Ram Murthy. C and Manoj B.S, "Ad Hoc Wireless Networks, Architectures and Protocols", Perason Edition Inc, 2005.

[2] Charles E. Perlins, "Ad Hoc Networking", Addision-Wesley 2nd Printing 2004.

[3] Elizabath. M and Belding. R, " Routing Approaches in Mobile Ad-Hoc Networks", 2004.

[4] Lei Wang, Lian Fang Zhang, Yantai Shu and Miao Dong. " Multipath Source Routing in Wireless Ad Hoc Networks", CCECE 2000, 7-10 March 2000.

[5] Mahesh. K and Marina S.R, "On-demand Multipath Distance Vector Routing in Adhoc Networks", 9th International Conference on 11-14 Nov 2001.

[6] Charles Perkins and Pravin Bhagwat, "Highly Dynamic Destination-Sequenced Distance-Ventor Routing for Mobile Computers", in Proceedings of the Symposium on Communication Architectures and Protocols, ACM SIGCOMM, 1994.

[7] Frederick Ducatelle, Gianni Di Caro and Luce Maria Gambardella, "Ant Agents for Hbride Multipath Routing in Mobile Ad Hoc Networks", Second Annuml conference on Wireless On-demand Network systems and services, 19-21 Jan 2005, WONS 2005.

[8] A.Iwata, Chiang C.C, Pei.G, Gerla.M and Chen T.W, "Scalable Routing Strategies for Ad Hoc Wireless Networks", IEEE Journal on Selected Areas in Communications, Vol 17, No.8, pp 1369-1379, August 1999.

[9] Chen.T.W and Gerla.M " Global State Routing: A New Routing Scheme for AdHoc Wireless Networks", Proceedings of IEEE, ICC 1998, pp 171-175, June 1998.

[10] Y.Ko and N.H,Vaidya, "Location-Aided Routing (LAR) in Mobile AdHoc Networks", Proceedings of ACM MOBICOM 1998, pp 66-75, October 1998.

[11] Soumendra Nanda, Robert S.Gray, "Spatial Multipath Location Aided AdHoc Routing", Computer Communications and Networks, 2004. ICCCN 2004 Proceedings, 13th International Conference.

[12] Liao W.H, Tseng Y.C and Sheu J.P, "GRID: A Fully Location-Aware Routing Protocol for Mobile AdHoc Networks", Telecommunication Systems, Vol 18, No.1, pp 37-60, 2001.

[13] Singh.S, Woo.M and Ragavendra. C.S, "Power-Aware Routing in Mobile AdHoc Networks", Proceedings of ACM MOBICOM 1998, PP 181-190, October 1998.

[14] Sushant Jain, "Energy Aware communication in AdHoc Networks", Technical Report, UW-CSE 03-06-03.

[15] Anand Prabhu Subramanian, Anto A.J, Janani Vasudevan and Narayanasamy .P, "Multipath Power Sensitive Routing Protocol for Mobile AdHoc Networks", WONS 2004, Madonna di Campiglio, Italy, Jan 2004.

[16] Sinha.P, Sivakumar.R and Bharghavan.V, "MCEDAR: Multicast Core Extruction Distributed AdHoc Routing", Proceeding of IEEE WCNC 1999, pp 1313-1317, Sep 1999.

[17] Bommaiah.E, Lin.M McAuley.A and Talpade.R, "AMRoute: AdHoc Multicast Routing Protocol", Internet Draft, draft-talpade-manet-amroute-oo.txt, August 1998.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[18] Shiwen Mao, Dennis Bushmitch, Sathya Narayanan and Shivendra S. Panwar, "MRTP: A Multi-Flow Real Time Transport Protocol for Adhoc Networks", Vehicular Technology Conference, 2003,6-9 Oct 2003.

[19] Ramon Fabregat, Yezid Donoso Meisel, Jose L.Marzo, Alfonso Ariza, "A Multi-Objective Multipath Routing Algorithm for Multicast Flows", 2004 International Symposium on performance Evaluation of Computer and Telecommunication Systems (SPECTS' 04).

[20] Souheila Bouam, Jalel Ben-Othman, "Data Security in Adhoc Networks using Multipath Routing", Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003, 14th IEEE Proceedings on Volume 2, 7-10 September 2003.

[21] Zhi Lij Yu-Kwong Kwok, " A New Multipath Routing Approach to Enhancing TCP Security in AdHoc wireless Networks", Parallel Processing, 2005. ICPP 2005 workshops, International Conference Workshops on 14-17, June 2005.

AUTHORS PROFILE

Udayachandran Ramasamy,(M'79,SM'94) born on 26.05.1956, completed his B.E (Electronics and Communication Engineering) in 1979, and M.E (Applied Electronics) in 1981 from P.S.G College of Technology, Coimbatore under University of Madras. Currently he is pursuing Ph.D at Vinayaka Missions University, Salem, India.

He has worked and carried out research in various institutions like REC, Tirchy, BIT, Bangalore, IOC, Assam, CIT, Coimbatore, VLBJACET, Coimbatore, Technical University of Novascotia, Canada Halifax, Concordia University, Montreal, Canada Institute Megatech, Kuala Lumput , Multimedia University, Kuala Lumpur.

He has 29 years of teaching and research experience. His area of interest includes Mobile AdHoc Networks Distributed Computing and Grid Computing. At present he is working as Professor and head, Department of Computer Science and Engineering, Sri Ramakrishna Institute of Technology, Coimbatore, Tamil Nadu, India. He is a Fellow of Institution of Engineers(India), Fellow Institution of Electronics and Telecommunication Engineers(India) and Senior Member IEEE(USA.)

K.Sankaranarayanan, born on 15.06.1952, completed his B.E (Electronics and Communication Engineering) in 1975, and M.E (Applied Electronics) in 1978 from P.S.G College of Technology, Coimbatore under University of Madras. He did his Ph.D (Biomedical Digital Signal Processing and Medical Expert System) in 1996 from P.S.G College of Technology, Coimbatore under Bharathiar University.

His areas of interest include Digital Signal Processing, Computer Networking, Network Security, Biomedical Electronics, Neural Networks and their applications, and Opto Electronics. He has more than 30 years of teaching experience and worked in various Government and Self financing Engineering Colleges.

At present he is working as PRINCIPAL at AKSHAYA College of Engineering and Technology, Coimbatore, Tamil Nadu, India.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Density Based Clustering Algorithm using Sparse Memory Mapped File

J. Hencil Peter
Department of Computer Science
St. Xavier's College, Palayamkottai , India.
hencilpeter@hotmail.com

A. Antonysamy
Department of Mathematics
St. Xavier's College, Kathmandu, Nepal.
fr_antonysamy@hotmail.com

Abstract:

The DBSCAN [1] algorithm is a popular algorithm in Data Mining field as it has the ability to mine the noiseless arbitrary shape Clusters in an elegant way. As the original DBSCAN algorithm uses the distance measures to compute the distance between objects, it consumes so much processing time and it's computation complexity comes as $O(N^2)$. In this paper we have proposed a new algorithm for mining the density based clusters using Sparse Memory Mapped File (Spares MMF) [3]. All the given objects are initially loaded into their corresponding Sparse Memory Mapped File's locations and during the SparseMemoryRegionQuery operation each objects' surrounding cells will be visited for the neighbour objects instead of computing the distance between each of the objects in the data set. Using the Sparse MMF approach, it is proved that the DBSCAN algorithm can process huge amount of objects without having any runtime issues and the new algorithm's performance analysis shows that proposed solution is super fast than the existing algorithm.

Keywords: Sparse Memory Mapped File; Sparse MMF; Sparse Memory; Neighbour Cells; Sparse Memory DBSCAN.

I. INTRODUCTION

Data mining is a fast growing field in which clustering plays a very important role. Clustering is the process of grouping a set of physical or abstract objects into classes of similar objects [2]. Among the many algorithms proposed in the clustering field, DBSCAN is one of the most popular algorithms due to its high quality of noiseless output clusters.

The most of the Density Based Clustering algorithms requires O (N^2) computation time and requires huge amount of main memory to process in the real time scenario. Since the seed object list grows during run time, it is very difficult to predict the required memory to process the entire objects present in the data set. If the memory is insufficient to process the growing seed objects, the DBSCAN algorithm will crash in the run time. So to get rid of the instability problem and improve the performance, a new solution has been proposed in this paper.

Rest of the paper is organised as follows. Section 2 gives the brief history about the related works in the same area. Section 3 gives the introduction of original DBSCAN and section 4 explains the proposed solution. After the new algorithm's explanation, section 5 shows the Experimental Results and final section 6 presents the conclusion and future work associated with this algorithm.

II RELATED WORK

The DBSCAN (Density Based Spatial Clustering of Application with Noise) [1] is the basic clustering algorithm to mine the clusters based on objects density. In this algorithm, first the number of objects present within the neighbour region (Eps) is computed. If the neighbour objects count is below the given threshold value, the object will be marked as NOISE. Otherwise the new cluster will be formed from the core object by finding the group of density connected objects that are maximal w.r.t density-reachability.

The OPTICS [4] algorithm adopts the original DBSCAN algorithm to deal with variance density clusters. This algorithm computes an ordering of the objects based on the reachability distance for representing the intrinsic hierarchical clustering structure. The Valleys in the plot indicate the clusters. But the input parameters ξ is critical for identifying the valleys as ξ clusters.

The DENCLUE [5] algorithm uses kernel density estimation. The result of density function gives the local density maxima value and this local density value is used to form the clusters. If the local density value is very small, the objects of clusters will be discarded as NOISE.

A Fast DBSCAN (FDBSCAN) Algorithm[6] has been invented to improve the speed of the original DBSCAN algorithm and the performance improvement has been achieved through considering only few selected representative objects belongs inside a core object's neighbour region as seed objects for the further expansion. Hence this algorithm is faster than the basic version of DBSCAN algorithm and suffers with the loss of result accuracy.

The MEDBSCAN [7] algorithm has been proposed recently to improve the performance of DBSCAN algorithm, at the same time without loosing the result accuracy. In this algorithm totally three queues have been used, the first queue will store the neighbours of the core object which belong

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

inside Eps distance, the second queue is used to store the neighbours of the core object which belong inside 2*Eps distance and the third queue is the seeds queue which store the unhandled objects for further expansion. This algorithm guarantees some notable performance improvement if Eps value is not very sensitive.

Though the DBSCAN algorithm's complexity can be reduced to O(N * logN) using some spatial trees, it is an extra effort to construct, organize the tree and the tree requires an additional memory to hold the objects. In this new algorithm different new complexity O ($N * 2^{Eps}$) has been achieved and it is proved that the new complexity better than the previous version of DBSCAN algorithms when the Eps value is minimal.

II. Introduction to DBSCAN Algorithm

The working principles of the DBSCAN algorithm are based on the following definitions:

Definition 1: Eps Neighbourhood of an object p

The Eps Neighbourhood of an object p is referred as NEps(p), defined as

$NEps(p) = \{q \in D \mid dist(p,q) <= Eps\}$.

Definition 2: Core Object Condition

An Object p is referred as core object, if the neighbour objects count >= given threshold value (MinObjs). i.e.

|NEps(p)|>=MinObjs

Where MinObjs refers the minimum number of neighbour objects to satisfy the core object condition. In the above case, if p has neighbours which are exist within the Eps radius count is >= MinObjs, p can be referred as core object.

Definition 3: Directly Density Reachable Object

An Object p is referred as directly density reachable from another object q w.r.t Eps and MinObjs if

$p \in NEps(q)$ and

|NEps(q)|>= MinObjs (Core Object condition)

Definition 4: Density Reachable Object

An object p is referred as density reachable from another object q w.r.t Eps and MinObjs if there is a chain of objects p1,…,pn, p1=q, pn=p such that pi+1 is directly density reachable from pi.

Definition 5: Density connected object

An Object p is density connected to another object q if there is an object o such that both, p and q are density reachable from o w.r.t Eps and MinObjs.

Definition 6: Cluster

A Cluster C is a non-empty subset of a Database D w.r.t Eps and MinObjs which satisfying the following conditions.

For every p and q, if $p \in$ cluster C and q is density reachable from p w.r.t Eps and MinObjs then $q \in C$.
For every p and q, $q \in C$; p is density connected to q w.r.t Eps and MinObjs.

Definition 7: Noise

An object which doesn't belong to any cluster is called noise.

The DBSCAN algorithm finds the Eps Neighbourhood of each object in a Database during the clustering process. Before the cluster expansion, if the algorithm finds any non core object, it will be marked as NOISE. With a core object, algorithm initiate a cluster and surrounding objects will be added into the queue for the further expansion. Each queue objects will be popped out and find the Eps neighbour objects for the popped out object. When the new object is a core object, all its neighbour objects will be assigned with the current cluster id and its unprocessed neighbour objects will be pushed into queue for further processing. This process will be repeated until there is no object in the queue for the further processing.

IV. Proposed Solution

A new algorithm has been proposed in this paper to improve the performance as well as to process huge amount of data. This algorithm is totally relying on Sparse MMF and the Sparse MMF concept has been explained below briefly:

A. Sparse Memory Mapped File (Sparse MMF)

The Sparse MMF [3] is the derived mechanism of Memory Mapped File. The Memory Mapped File [3] is like virtual memory and it allows reserving a region of address space and committing physical storage to the region. The difference is that the physical storage comes from a file that is already on the disk instead of the system's paging file. The memory mapped file can be used to access the data file on disk (even very huge files), load and execute executable files and libraries and allowing multiple processes running on the same machine to share data with each other. The Sparse MMF is similar to Memory Mapped File but it occupies only the required storage space in the physical file. If we use Memory Mapped File to reserve the region of memory, while committing the changes to the file on disk, the file size will be equivalent of the created Memory Mapped File size. Instead if we replace the same with Sparse MMF, final file's size will be equivalent to the e non-zero element which is

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

stored in the Sparse MMF. So Sparse MMF gives better storage result and hence it has been used in our research.

B. Object's Structure

As this algorithm's core is Spare MMF, the objects that needs to be processed by this algorithm are organized bit differently and each objects' structure will have three additional fields NextObjectOffset, NextSeedObjectOffset and NextTempObjectOffset.

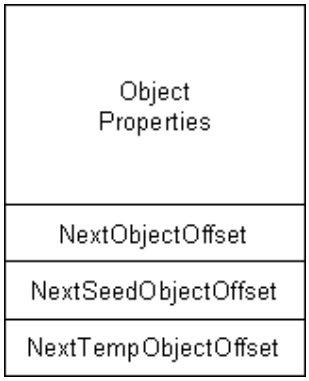


Figure 1. Sparse Memory Mapped File Object's Structure

While loading all the objects in Sparse MMF, all the objects are chained in a sequence like linked list (but not exactly linked list). The first additional field NextObjectOffset will hold the Offset value of the next object, second object will hold the offset of its immediate successor object, etc and the final object's NextObjectOffset will set to NULL to indicate that there are no more objects further to visit during the clustering process. So the first object's address should be retained always to visit the entire objects loaded in the Sparse MMF. The other two fields NextSeedObjectOffset and NextTempObjectOffset fields are used by SparseMemoryRegionQuery function call and it is explained in the below section.

C. SparseMemoryRegionQuery function

The proposed algorithm doesn't uses any extra buffer or queue to store the seed objects as well as neighbour objects during the run time, instead each object has the corresponding Offset field and in which the exact offset of the next seed object will be stored. In the original DBSCAN algorithm, RegionQuery function has been used to retrieve the neighbour objects and in this new algorithm SpareMemoryRegionQuery function has been introduced instead of RegionQuery. This function visits all the required surrounding cells in memory and the non empty cell objects will be chained and return back as seed objects. i. e The function start from the center cell and visit the neighbour cells one by one. When the non empty object found in the first time, center object's NextSeedOffset field will be assigned the Offset of new object (Address(NewObject) – Address(CenterObject)) and next time when the new object found, current object's offset will be stored in the previous object's NextSeedObject field and so on. Eventually last object's NextSeedObject field will be assigned with NULL. Thus the extra memory as well as buffer/queue requirement to store the seed objects has been removed in this solution. This function has been customized to update the neighbour objects offset in the either field NextSeedObectOffset or NextTempObjectOffset. If this function receives an update flag UpdateMasterSeedOffset, neighbour objects offset will be stored in NextSeedObectOffset field and input update flag is UpdateTempSeedOffset then the NextTempObjectOffset will be updated with the neighbour object(s) offset.

The DBSCAN algorithm's computation complexity varies based on the RegionQuery function and it uses distance function to compute the neighbours present with in the certain radius (Eps). In this new approach, distance computation during the SparseMemoryRegionQuery function call has been removed and it visit's the required number of neighbour cells from the center cell.

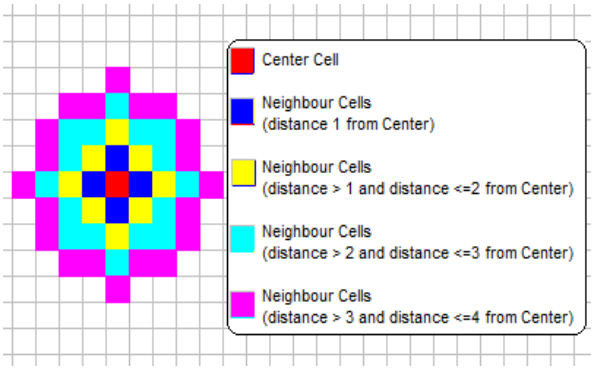


Figure 2. Neighbour Cells Diagram

In this proposed solution, we have selected two dimensional dataset for the experiment and the above diagram shows the neighbour cells with different distance. The center cell has been painted in red colour and it's distance of object stored in the cell will be zero, next immediate neighbours whose distance is 1 from the center cell have been painted in blue colour, the yellow colour cells distance are greater than 1 and <=2 and so on. These neighbour cells offsets are pre-computed and stored in M X 2 dimensional array and it will be passed to the SparseMemoryRegionQuery function to visit only the required number of neighbour cells to process. Thus the distance computation between objects is not required.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

D. Neighbour Cells and Index Offset Array

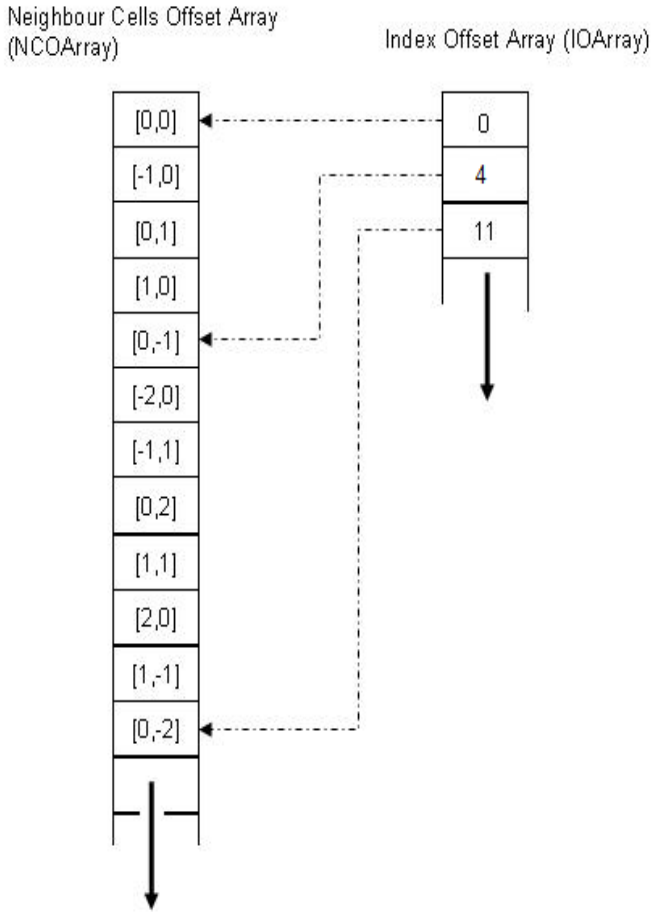


Figure 3. NCOArray and IOArray

Two additional arrays are been used in this algorithm to avoid the distance computation and improve the performance. The first array Neighbour Cells Offset Array (NCOArray) is an M X 2 array and it stores the offset values of neighbour cells from the center object. The Second Index Offset Array (IOArray) is K X 1 dimensional array and it stores the NCOArray's last index value for the corresponding Eps value sequence starting from 0. For example if the Eps value is 1 then IOArray[1] tells that NCOArray array elements starting from 0 to 4 have the cells offset that need to be visited by SparseMemoryRegionQuery during the neighbour objects computation. The value M will be decided based on the maximum possible Eps value supported by the algorithm and based on this K value also determined. So these two array values are populated with the required values before the actual clustering process.

E. Algorithm

1) Input D, Eps, MinObjs.
2) Create SparseMemoryMapped File.
3) Load the pre-computed Neighbour Cells Offset Array "NCOArray" and Offset Index Array "OIArray" Values.
4) Initialize the SparseMemoryMapped file with the dataset D, assign ClusterID field of all objects with UNCLASSIFIED and preserve the First Object's Address.
5) ClusterID = NOISE, CurrentObject = FirstObject.
6) WHILE CurrentObject <> NULL
7) If (CurrentObject.ClusterID == UNCLASSIFIED) Then
8) Call SparseMemoryRegionQuery function with CurrentObject, Eps, UpdateMasterSeedOffset, NCOArray and OIArray parameter and the function returns FirstSeedObject, LastSeedObject and SeedObjectsCount.
9) If (SeedObjectsCount >= MinObjs) Then// Core Object condition
10) ClusterID = GetNextID(ClusterID).
11) Assign the ClusterID to all the seed objects.
12) Move CurrentSeedObject to point its next seed object using the OffsetValue and assign NULL value to previous CurrentSeedObject's NextSeedObjectOffset field.
13) WHILE CurrentSeedObject <> NULL
14) Call SparseMemoryRegionQuery function with CurrentSeedObject, Eps, UpdateTempSeedOffset, NCOArray and OIArray parameter and the function returns TempFirstSeedObject, TempLastSeedObject and TempSeedObjectsCount.
15) If (TempSeedObjectsCount >= MinObjs) Then
16) TempCurrentSeedObject = TempFirstSeedObject.
17) For I = 1 to TempSeedObjectsCount
18) If TempCurrentSeedObject .ClusterID IN {UNCLASSIFIED, NOISE} Then
19) If TempCurrentSeedObject.ClusterID == UNCLASSIFIED Then
20) Append the TempCurrentSeedObject to the LastSeedObject.
21) End If
22) TempCurrentSeedObject .ClusterID = ClulsterID.
23) End If
24) Move TempCurrentSeedObject to point its next seed object using the OffsetValue and assign NULL value to previous TempCurrentSeedObject's NextTempSeedObjectOffset field.
25) End For
26) End If
27) If (CurrentSeedObject. NextObjectOffset == 0)

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Then
28) CurrentSeedObject =NULL.
29) Else
30) Move CurrentSeedObject to point its next seed object using the OffsetValue and assign NULL value to previous CurrentSeedObject's NextSeedObjectOffset field.
31) End If
32) END WHILE
33) Else //Non Core Object
34) CurrentObject.ClusterID = NOISE.
35) Assign NULL value to all the SeedObjects' NextSeedOffset member.
36) End If
37) End If
38) If (CurrentObject. NextObjectOffset == 0) Then
39) CurrentObject=NULL.
40) Else
41) Move CurrentObject to point its next object using the OffsetValue.
42) End If
43) END WHILE

This algorithm starts with creating the Sparse MMF with the required size and loads the Neighbour Cell Offset and Index Offset array values. The dataset D will be read one by one and each object will be placed in the corresponding memory locations. As mentioned in the section 4(B), while initializing the Sparse MMF with objects, each successive object's memory offset will be stored in the previous objects NextObjectOffset field and last object's NextObjectOffset field will be assigned with NULL value. Thus it is very essential to preserve the FirstObject's address to visit all the remaining objects.

The algorithm starts the traverse from the first object and visits the next objects one by one using the next object's offset stored in the current object itself. When it finds the object and its cluster ID is UNCLASSIFIED, SparseMemoryRegionQuery function will be called with required parameter. As the new cluster is not yet formed, SparseMemoryRegionQuery function needs to be called with UpdateMasterSeedField flag to update the seed objects' NextObjectSeedOffset field. The output of SparseMemoryRegionQuery will give FirstSeedObject, LastSeedObject and SeedObjectsCount. If the current object is a non core object, the current object will be market as NOISE and all its seed objects NextObjectSeedOffset field will be market with NULL value. Otherwise the cluster expansion will start with creating a new cluster ID as the current object is a core object. The new Cluster ID will be assigned to all the seed objects that are chained starting from FirstSeedObject. Now the remaining objects (except FirstSeedObject) present in the seed chain will be processed one by one and for all the remaining seed objects SparseMemoryRegionQuery will be called with UpdateTempSeedOffset flag to update the TempObjectNextSeedOffset field. This will avoid the overwriting of seed objects which are already exist in the main seed list chain. So if the object is a core object, the neighbour objects will be visited using TempObjectNextSeedOffset instead of ObjectNextSeedOffset and the UNCLASSIFIED cluster id type objects present in the temporary seed chain will be appended to the LastSeedObject (main seed chain) for the further processing and all the UNCLASSIFIED and NOISE type objects present in the temporary seed list will be assigned with the current Cluster ID. The LastSeedObject member will always point the last object in the seed chain. The entire object present in the main seed chain will be processed one by one and cluster expansion will stop when the traverse reaches the LastSeedObject and no more seed objects to process further. The complete clustering process will stop once the initial loop process the entire objects present in the data set.

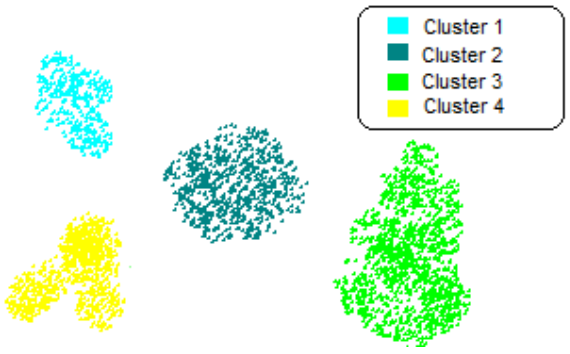


Figure 4. Result of Dataset 1

F. Advantages

The proposed algorithm is very stable. The main drawback of original DBSCAN algorithm is instability. Though all the objects present in the data set can be loaded by the DBSCAN algorithm, if we don't have sufficient main memory to hold the growing seeds objects, DBSCAN algorithm will crash during run time. But the new algorithm doesn't rely on the growing seeds and it will give guarantee to process all the objects as long as it is able to load. The second advantage of the new algorithm is capable of processing huge amount of objects. Since this algorithm is based on the Sparse MMF, it can support few GBs of data in a 32 bit Operating System where traditional approach supports only few MBs of data in the real time scenario. Also this algorithm can be customised to process very huge data set (e.g > 10 GB) using the Sparse MMF. Then the beauty of Sparse MMF is, though we pre-allocate more memory in the beginning, the real memory occupying is based on the consumption. Eventually the performance is really fast as the algorithm directly works on the memory.

G. Limitations

As this algorithm uses Sparse MMF and only very few languages support this feature, scope for implementing this algorithm is limited. Second limitation is memory

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

customization. If we are planning to apply this algorithm to support multidimensional dataset, memory needs to be customized accordingly and the computation complexity may vary. Also if the minimum distance between one object and the immediate nearest object is greater than one unit or less than one unit, offset array values will change and it should be recomputed. Moreover creating and populating values in Offset arrays are an extra task. Last drawback of this algorithm is this doesn't support duplicate objects. As the object loaded in the corresponding memory location, it is not possible to overwrite another object in the same location. These are the notable limitations of this algorithm.

H. Computation Complexity

The DBSCAN algorithm's complexity has been calculated based on the number of RegionQuery function call. In which each RegionQuery function call need N distance computation and hence the computation complexity becomes O (N^2) for processing all the N objects present in the dataset. As the new algorithm's SparseRegionQuery process the neighbour cells, the complexity varies based on the Eps value and each SparseRegionQuery requires not more than $2^{(Eps+1)}$ cells traversal. Eventually for processing all the N objects, our algorithm requires O ($N * 2^{(Eps+1)}$) time. The constant 1 can be removed as it is very small and the final complexity comes as O ($N * 2^{Eps}$). This complexity is really a reduction when the Eps value is reasonable (e.g 1~10) and N value is very large. At the same time, if we have very less number of objects and the Eps value is too big, this new complexity won't be an attractive one. However the real processing time will be very faster than the traditional RegionQuery function call as the SparseRegionQuery traverse the memory directly.

TABLE 1. COMPARISON OF ALGORITHMS

Algorithm	Stability (because of growing Seed)?	Doesn't Require extra Buffer	Ability to process huge dataset?	Supports Duplicate Objects?	Better Performance	Doesn't depend on distance function.
DBSCAN	No	No	No	Yes	No	No
DBSCANSMMF	Yes	Yes	Yes	No	Yes	Yes

Above table show the comparison of some key features and DBSCANSMMF is superior in most of the features.

V. EXPERIMENTAL RESULTS

The newly proposed algorithm and the original DBSCAN algorithm have been implemented in Visual C++ (2008) on Windows Vista OS and ran on PC with a 2.0 GHZ processor and 4 GB RAM to observe the performance. The different sizes of 2 dimensional synthetic datasets were used and running time results are given below:

TABLE 2. RUNNING TIME OF DBSCAN AND DBSCANSMMF IN SECONDS

Number of Objects	1.DBSCANSMMF	1.DBSCAN	2.DBSCANSMMF	2.DBSCAN
1500	0.0007	0.3892	0.0005	0.2176
3000	0.0043	0.5395	0.0051	0.5684
6000	0.0081	1.8030	0.0094	1.8920
10000	0.0137	4.9124	0.0166	5.1122
20000	0.0261	20.4426	0.0255	18.2351
30000	0.0377	43.3875	0.0269	41.1765
40000	0.0545	77.6204	0.0587	79.6543
60000	0.0799	195.8284	0.0676	181.8745

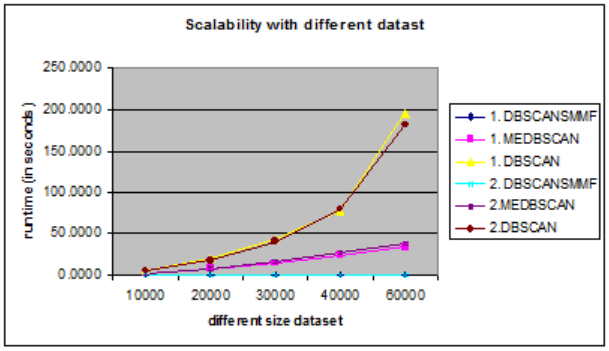


Fig 5. Scalability of Algorithm with different size of dataset

The above table and graph figures show that new algorithm gives better performance when the algorithm's input data set size grows. This is the expected obvious result as the new algorithm visits only the required neighbour cells during the SparseMemoryRegionQuery function call instead of the computing distance between center and the entire objects in the data set. Another reason is directly accessing the memory is much faster than using the buffers to process the data that are usually used to implement the algorithms.

VI. CONCLUSION AND FUTURE ENHANCEMENT

In this paper we have proposed DBSCANSMMF algorithm to improve the performance as well as to process the huge amount of data using Sparse MMF. This new algorithm doesn't uses any growing seed list which causes the crash during the run time when there is no sufficient memory to store the seed objects. Instead the new algorithm just maintains the seed list using the offset values and these values are stored in each objects corresponding offset field internally. So there is no need of creating duplicate objects

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

for processing the objects. Also this new algorithm takes O ($N * 2^{Eps}$) computation complexity and this is better complexity as long as Eps value is reasonable.

Future work will be to customize this algorithm to support duplicate objects. This can be achieved using the internal counter which will give the number of similar objects and the SparseMemoryRegionQuery also needs to be customized accordingly to support correct output. The next expansion will be customizing this algorithm to process super big data set (e.g. 50 GB). One of the real uses of Memory Mapped File is mapping the required portion of the file into memory to process and, un map the current mapped region and remap the next consecutive file region to process later. Like this we can process any big file and this algorithm needs to be customized to support this feature.

REFERENCES

[1] Ester M., Kriegel H.-P., Sander J., and Xu X. (1996) "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise" In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD'96), Portland: Oregon, pp. 226-231

[2] J. Han and M. Kamber, Data Mining Concepts and Techniques. Morgan Kaufman, 2006.

[3] Jeffrey Richter and Christophe Nasarre, WINDOWS VIA C/C++, Microsoft Press, 2008.

[4]M. Ankerst, M. Breunig, H. P. Kriegel, and J. Sander, "OPTICS: Ordering Objects to Identify the Clustering Structure, Proc. ACM SIGMOD," in International Conference on Management of Data, 1999, pp. 49–60.

[5] A. Hinneburg and D. Keim, "An efficient approach to clustering in large multimedia data sets with noise," in 4th International Conference on Knowledge Discovery and Data Mining, 1998, pp. 58–65.

[6]SHOU Shui-geng, ZHOU Ao-ying JIN Wen, FAN Ye and QIAN Wei-ning.(2000)
"A Fast DBSCAN Algorithm" Journal of Software: 735-744

[7] Li Jian; Yu Wei; Yan Bao-Ping; , "Memory effect in DBSCAN algorithm," Computer Science & Education, 2009. ICCSE '09. 4th International Conference on , vol., no., pp.31-36, 25-28 July 2009.

AUTHOR PROFILES

J. Hencil Peter is Research Scholar, St. Xavier's College (Autonomous), Palayamkottai, Tirunelveli, India. He earned his MCA (Master of Computer Applications) degree from Manonmaniam Sundaranar University, Tirunelveli. Now he is doing Ph.D in Computer Applications and Mathematics (Interdisciplinary) at Manonmaniam Sundranar University, Tirunelveli. His interested research area is algorithms inventions in data mining.

Email: hencilpeter@hotmail.com

Dr.A. Antonysamy is Principal of St. Xavier's College, Kathmandu, Nepal. He completed his Ph.D in Mathematics for the research on "An algorithmic study of some classes of intersection graphs". He has guided and guiding many research students in Computer Science and Mathematics. He has published many research papers in national and international journals. He has organized Seminars and Conferences in state and national level.

Email: **fr_antonysamy@hotmail.com**.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Mapping Cloud Computing onto Useful e-Governance

Ajay Prasad
Department of CSE
Sir Padampat
Singhania University, Udaipur,
INDIA
ajayprasadv@gmail.com

Sandeep Chaurasia
Department of CSE
Sir Padampat
Singhania University, Udaipur,
INDIA
Sandeep.chaurasia@spsu.ac.in

Arjun Singh
Department of CSE
Sir Padampat
Singhania University, Udaipur,
INDIA
vitarjun@gmail.com

Deepak Gour
Department of CSE
Sir Padampat
Singhania University, Udaipur,
INDIA
deepak.gour@spsu.ac.in

***Abstract*— Most of the services viewed in context to grid and cloud computing are mostly confined to services that are available for intellectual purposes. The grid or cloud computing are large scale distributed systems. The essence of large scale distribution can only be realized if the services are rendered to common man. The only organization which has exposure to almost every single resident is the respective governments in every country. As the size of population increases so the need for a larger purview arises. The problem of having a large purview can be solved by means of large scale grid for online services. The government services can be rendered through fully customized Service-oriented Clouds. In this paper we are presenting tight similarities between generic government functioning and the service oriented grid/cloud approach. Also, we will discuss the major issues in establishing services oriented grids for governmental organization.**

Keywords- Grid computing, cloud computing, service oriented grids, e-governance, SOA, CMMS..

I. Introduction

The analogy of electricity grid in compute grids is appropriately understood by the fact that both grids are mainly meant for supplying vastly distributed resources and services. The figure 1 depicts the exact similarities between them. Virtualization is the main approach towards grids.

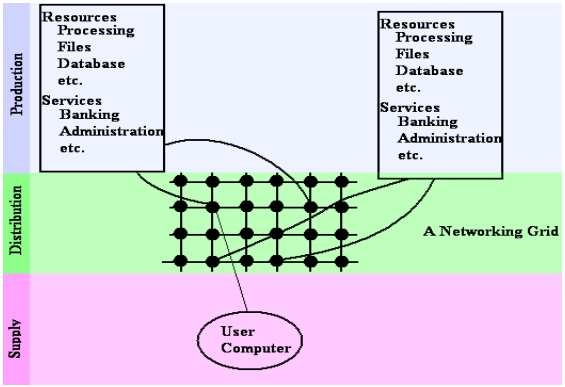


Figure 1. A Compute Grid

The Virtual organization (VO) based grids are those where services fall out of organizational boundaries, and multiple organizations combined render services to users. The user connects to a discovery node nearby it and gets to a service node in order to get service. Figure 2 shows a cloud containing a set of valid nodes ready to provide service. The service oriented grid middleware for e-learning as suggested in [1] needs to be furthered in aspects of e-governance.

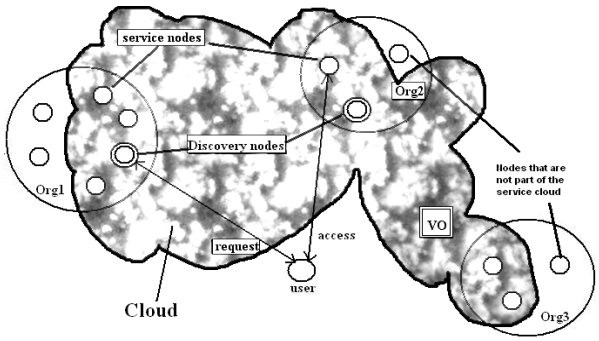


Figure 2. A Virtual Organization involving many organizations together (cloud).

Though many governments have already initiated in the aspects of computerization of certain services but the need of the time is to have an integrated approach to all types of government services under the framework of more customizable, distributed and scalable system, that is, obviously grids or clouds.

II. Government services and service oriented grid / cloud approach

As depicted in the figure 3 almost all government services functionally consists of 4 layers as far as manual systems are concerned. Without applying much change in the traditional levels of operation, a grid system can be set up which adds up more features like quick processing, transparency, reliable and above all accessible to common man.

The political or regulatory body consists of the state and the highest level officials (mostly politicians). The enforcement body will cover all districts and will consist of officials appointed by the state. The servicing body will cover divisions as-well as sub divisions. This body will be the actual face of the government operations and will interact with the masses directly.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The users are the citizen of the state viewed as part of a particular division or sub division. The grid system if incorporated will deem both government officials and citizen as its users as all of them will be using one or other services / applications available on the grid system (not the manual system). Only the authorization levels or state [8] will be different. An example in this regard is well depicted in [8] where a prototype grid on CMMS authorization model is presented for a state police services.

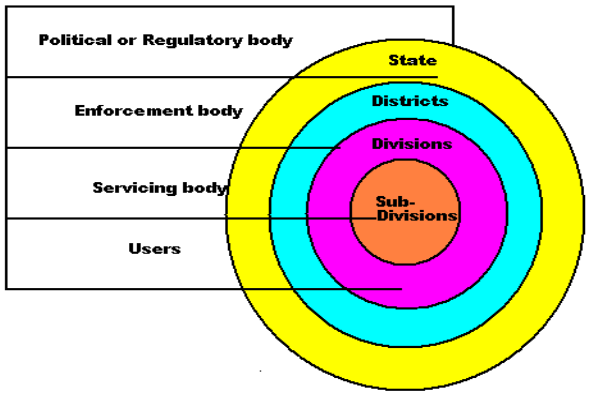


Figure 3. Generic functional structure of any government service.

The Grid or Cloud will consist of major 3 types of nodes namely:

1. Monitoring node.
2. Controlling node.
3. Service node.

The figures 4, 5 depict a functional structure of the grid/cloud. The functionality of monitoring node might be that of assigning roles, changing states [8], changing/creating/approving certificates etc. The monitoring node will have the major task of acting as discovery nodes, establish synchronization among service nodes, managing scaling of service nodes, imposing states [8] etc. The service nodes will be the actual nodes at various sites which will be operated by the service delivery officials who will use the end software to update records and perform transactions. The users will be accessing the grid through peer grid softwares or may be via internet.

III. Infrastructural and Implementation Issues

We can broadly categorize the cloud e-governance implementation issues into following three factors:

A. Infrastructure or backbone development.
B. Software implementation.
C. Ensuring usability.

A. Infrastructure or backbone development

The grid/cloud based e-governance at its peak would require a well made and managed infrastructure throughout the province and beyond. The possibilities of having an eco-grid [2] for optimum resource utilization and at a economic mode can be framed similarly as in [2].

Structurally the clouds will consist of nodes servicing at remote frontiers. The users might well be accessing from long distances. The service nodes may need to synchronize in many cases and that too quite frequently. The controlling nodes of these service nodes might well be at some well defined establishments and may be rendering services along with control. The database synchronization and maintenance will be both at node level as well as grid/cloud level.

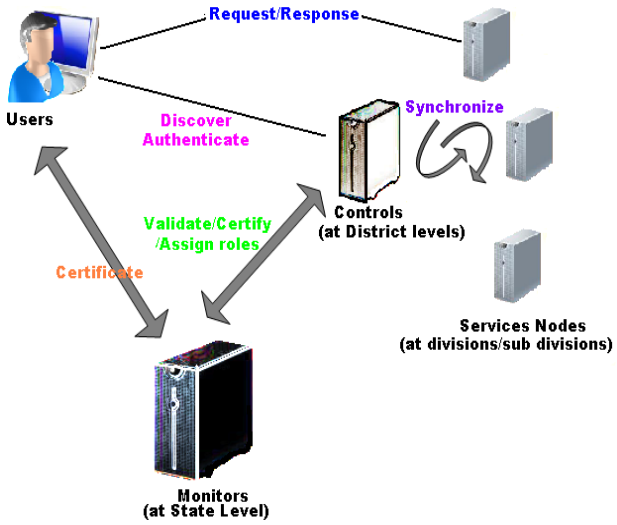


Figure 4. Role players in the Grid/Cloud.

B. Software implementation

Among many issues pertaining to any grid establishment, the factors of concurrency, security, and performance are major ones. The OGSI, WSRF, GSI [10] standards do take care of almost completely. The factor of authorization is vital in grid services and has to be taken care of separately than just depending on standards (GSI) alone. The CMMS model [8] suggests an authorization procedure especially for peer grid services. The model is mainly designed keeping in mind the governmental services.

Service oriented architecture (SOA): The e-governance should be implemented over SOA As rightfully acknowledged in [3], the SOA brings exceptionally component based development with Enterprise Application Integration (EAI). Also, the fact of having peer to peer SOA (SOPs [4]) is more suitable with the general model presented for authorization i.e. CMMS [8].

The establishing of the entire infrastructure for internet based clouds might require sincere and serious effort from the state but once achieved can make a comfortable environment of governance and helps the governments to have better understanding of the need and demand of the polity as well as the mass.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The State Wide Area Networks (SWAN) [12] plan with the Indian government is a vital step towards achieving fruitful e-governance. The infrastructure will help in providing a backbone for an e-governance cloud for efficient governance of the mass.

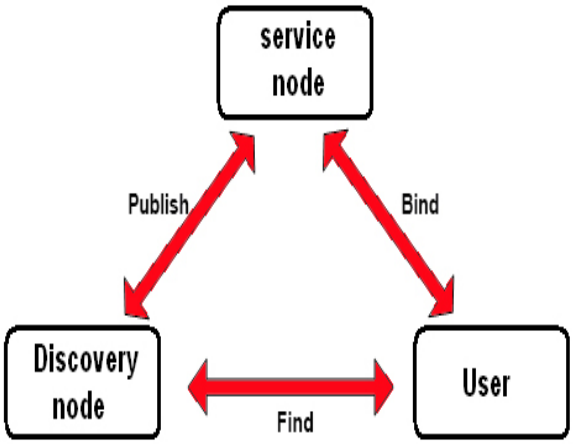


Figure 5. Basic service oriented architecture over grids.

Every grid applications has to rely upon a certification procedure that is robust, well defined and managed by a reliable third party. In case of using the CMMS model the third party will be the highest body in the state. However, since giving the keys in the hands of regulatory bodies may invite digital corruption, that is, corruption related to manipulation of digital signatures/certificates. Giving a secret and uniform identification numbers to all citizens would be required in order to have a corruption free and hassle free grid based e-governance application. The Nilekani's UID project [9, 11] in India is a step forward to realize the dream of services oriented grids for e-governance.

C. *Ensuring usability*

Ensuring usability is of prime concern as far as respective governments are concerned. A joint impact assessment study [5] was initiated in January 2006 by Indian Institute of Management and e-governance practice group of World Bank, Washington DC. The study was to define a framework and methodology for impact assessment of e-government projects from 3 states in India and 2 projects from Chile. The primary objective of this study was to measure the impact of computerization in selected service delivery projects. Though all the projects under assessment are currently for servicing urban clients, a promising fact comes out of this is, if the population is computer literate and has an infrastructural setup then gradually but steadily the manual delivery service is bound to get abolished.

It is quite certain that there is a huge mandate for the computerized systems. As for example KAVERI [4] was opened to service stamp duty and registration sector was started in December 2003. In 2000-01 when the system was manual only 0.63 million properties were registered. In 2005-06, with KAVERI system 1.02 million properties were registered representing an annual growth of 10.27%. Nearly 98% of respondents preferred the computerized system over the manual system because the time and cost of getting the service for clients has come down significantly. The survey also pointed out the major factors that the citizen demand for certain services. They prepared a chart by interviewing a set of citizen for their most desired factors for certain services like land registration, procurement, taxation etc. The table I is the original table revealing top four most desired attributes in particular software services. By having a closer look into the table we can easily figure out that major factors that are desired are improved governance, Quality and transaction efficiency in any services rendered through online or computerized system. Same are the prime factors must be taken care of on a servicing grid for e-government services. Among the three most desired attribute, the quality plays a major portion. Thus, the designers of the software over the grid infrastructure have to be careful that the quality factors are not compromised with.

Project	Attribute 1	Attribute 2	Attribute 3	Attribute 4
KAVERI	Less Corruption	Greater Transparency	*Error free transaction*	**Less waiting time**
Khajane – DDO	*Simplicity of procedure*	*Convenient time scheduled*	*Friendly attitude of officers*	*Error free transaction*
Khajane – Payee	**No delay in transaction**	*Convenient time scheduled*	*Good location*	*error free transaction*
eProcurement	No corruption	*Easy access*	*Equal opportunity to all*	*No need to visit Government Office*
eSeva	**Less time and effort required**	**Less waiting time**	*Convenient time scheduled*	*Equal opportunity to all*
Checkpost	**No delay in transaction**	*Error free receipt*	*Error free transaction*	*Proper queue system*
Legend: Underline - Improve Governance; Bold - Transaction Efficiency; Italics - Quality				

TABLE I. TOP FOUR ATTRIBUTES DESIRED IN CERTAIN COMPUTERIZED SERVICES SURVEYED IN INDIA.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

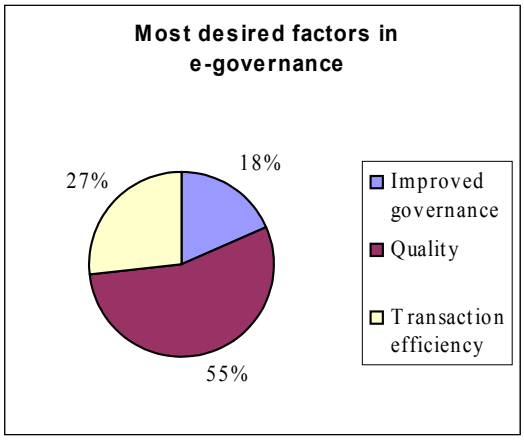


Figure 6. Percentage desire for generalized attributes in an e-governance software.

Major quality factors are: error free receipt, easy access, equal opportunity to all, proper queue system etc. Grid itself is a big solution for easy access. The certification authority and the authentication procedures have to be tight enough so that receipts are error free and can be regenerated if demanded.

In most of the cases the citizen users have always appreciated the online services and are sure to appreciate the fast, robust and reliable services on the grid infrastructures. Coming to verse with any new system is not difficult for any one provided a proper introduction and helping hand is extended. Nilekani [6] cites one instance after another of rural village dwellers, farmers, taxpayers, and others who quickly grasp what computers and Internet access can do for them. Whether it's the chance to learn English, check crop prices, or pay a utility bill, Indians at all levels come to depend on the computer once it's introduced. (The hard thing is persuading agencies and local officials to install systems that undercut their power as gatekeepers.) And we've all heard of the Hole in the Wall Project, where Indian kids in slums come to enjoy and figure out how to use computers with little or no adult help. This is true for any country and not only India.

IV. Conclusion

Service oriented grids/clouds are fast emerging these days. In this paper we discussed how a service oriented grid could be utilized to provide useful e-governance. Government services can become more reliable and transparent. Services oriented grids can be realized using the SOA and CMMS models. Few primary layouts were discussed by showing the similarities between government structures and the suggested CMMS model. The issues pertaining to the realization of grid based e-governance were discussed emphasizing the points like implementation, usability and infrastructure. Current governments should take due steps to build a favorable infrastructure for grids. The citizens have to be empowered by steps like UID etc. sustained efforts from both government and citizens could help build a realistic service oriented grid that could benefit even the most common citizen of the state. The paper is intended to bring more light into this direction so that more and more intellect could be driven towards it. More sustained efforts are required to develop a complete and more concrete design of the service oriented grids for useful e-governance.

V. References

[1] Wang GuiLing, Li YuShun, Yang ShengWen, Miao ChunYu, Xu Jun, Shi MeiLin, "Service-Oriented Grid Architecture and Middleware Technologies for Collaborative E-Learning," scc, vol. 2, pp.67-74, 2005 IEEE International Conference on Services Computing (SCC'05) Vol-2, 2005

[2] Rajkumar Buyya, David Abramson, Jonathan Giddy**, "**A Case for Economy Grid Architecture for Service Oriented Grid Computing", Proceedings of the 10th Heterogeneous Computing Workshop HCW 2001 (Workshop 1) - Volume 2 20083.1, 2001

[3] Alexander Sterff, "Analysis Of Service-Oriented Architectures From A Business And An It Perspective", Master's Thesis, 2006

[4] KAVERI - Karnataka Valuation And e-Registration , www.karigr.org

[5] Impact Assessment Study Of E-Government Projects In India, Center For e-Governance, Indian Institute Of Management, Ahmedabad, Department Of Information Technology, Government Of India, New Delhi,January 2007, http://www.iimahd.ernet.in/egov/documents/impact-assessment-study-dit.pdf

[6] Andy Oram, "Computerization in Nilekani's Imagining India", http://radar.oreilly.com/2009/09/computerization-in-nilekanis-i.html, as viewed on date 11-06-10

[7] "Structure of Indian Government" http://www.atatwork.org/content/download/427/2763/file/Attachment%2011%20Structure%20of%20Indian%20Government.pdf

[8] Ajay Prasad, A.K. Sharma, S.S. Verma, "Certification Authority Monitored Multilevel and Stateful Policy Based Authorization in Services Oriented Grids", International Journal of Security, (IJS), Volume (3) : Issue(4):48-64, 2009

[9] Creation of position of the Chairperson UID Authority of India, http://pib.nic.in/release/release.asp?relid=49370, as on 12.30 pm, 05/07/2010.

[10] Bart Jacob, Michael Brown, Kentaro Fukui, Nihar Trivedi. Introduction to Grid Computing, ibm.com/redbooksibm.com/redbooks. IBM Corporation 2005 Dec 2005.

[11] Nilekani takes charge, TNN, Jul 24, 2009, 05.06am IST, http://timesofindia.indiatimes.com/articleshow/4812763.cms, as on 12.30 pm, 05/07/2010

[12] State Wide Area Network (SWAN), http://www.nisg.org/kchome.php?page=9, as on 12.30 pm, 05/07/2010

Mr. Ajay Prasad, Assistant Professor, Dept of CSE, Sir Padampat Singhania University did his M.Tech. in CSE from MITS University, Laxmangarh, following an MCA Degree from Nagpur University. He did his bachelors (B.Sc.) from Osmania University, Hyderabad. He has been rendering his services in Academic as well as s/w development fields since last 12 years and has been mentoring lots of project works ranging from academics to real time. He has contributed in the fields of security in Grid/Cloud and adhoc networks and security in social networks and he has been researching in probabilistic approaches in data

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

mining techniques. He is Life member of ISTE. Currently he is pursuing his PhD from SPSU.

Mr. Sandeep Chaurasia, Lecturer, Dept of CSE, Sir Padampat Singhania University did his M.Tech. in CSE from Devi Ahilya University, Indore. He did his bachelors (B.E) from Rajasthan University, Jaipur. He has in academics and contributing as well in s/w development fields since last 3 years. He has contributed in the fields of application development in .NET Technology and software testing and he has been researching in Management Information Systems and knowledge management in business organization with quantification of knowledge management. Currently he is pursuing his PhD from SPSU.

Mr. Arjun Singh, Senior Lecturer in Sir Padampat Singhania University. He did his M.Tech Degree with specialization in networking from VIT Vellore. His specialization is mainly in the Network Security and cryptography, wireless Grid Computing, Computer Network, Network design. Prior to his teaching assignment, he served in IBM & Capgemini pvt. Ltd. as a software engineer.

Mr. Deepak Gour, Sr. Lecturer – Dept. of CSE, Sir Padampat Singhania University, Udaipur completed his B.Sc. in Computer Science in 1998 & Master in Computer Application in 2001. Currently he is Perusing Ph.D. from Department of Computer Science, Mohan Lal Sukhadia University, Udaipur. His research area is in ASIP Design Space Exploration. His Area of Specialization is in Embedded Systems and his Research interest lies in Application Specific Instruction set Processor.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Honey-Bee Mating based Bound Time Approach for Energy minimization in Wireless Sensor Networks

J Senthilkumar
Assistant Professor/IT
Sona College of Technology
Salem, Tamilnadu, INDIA
jsenthil10@gmail.com

Dr.R.Lakshmipathi
Professor/EEE
St.Peters Engineering College
Chennai, Tamilnadu, INDIA
drrlakshmipathi@yahoo.com

V.Mohanraj
Assistant Professor/IT
Sona College of Technology
Salem, Tamilnadu, INDIA
rajkrishcounty@gmail.com

Y.Suresh
Assistant Professor/IT
Sona College of Technology
Salem, Tamilnadu, INDIA
shuresh_22@yahoo.co.in

***Abstract*—In Wireless Sensor Network, dynamic cluster based routing protocol approach is widely used. Such practiced approach, quickly depletes the energy of cluster heads and induces the execution of frequent re-election algorithm. This repeated cluster head re-election algorithm increases the number of advertisement messages which in turn depletes the energy of overall sensor network. Here, we proposed the Bound Time and Honey Bee Mating Approach (BT-HBMA) that reduces the cluster set up communication overhead and elects the stand by node in advance for current cluster head which has the capability to withstand for many rounds**

Our proposed BT-HBMA method uses the Honey bee mating behaviour in electing the stand by node for current cluster head. This approach really outperforms the other methods in achieving reduced number of re election and maintaining high energy nodes between the rounds.

Keywords- Cluster based Routing; Wireless sensor network; Honey Bee mating; Bound Time

I. INTRODUCTION

Wireless Sensor Networks (WSNs) are formed by a set of nodes that gather information and forward it to a sink. They are formed by small, inexpensive and resource limited devices that can interact with the environment and communicate in a wireless manner with other devices [1] WSNs present a new challenge research problem due to their high flexibility to support several real-world applications. The core operation of wireless sensor network is to collect and process data at the network nodes, and transmit the necessary data to the base station for further analysis and processing. Due to large network size, limited power supply, and inaccessible remote environment, the WSN-based protocols are different from the traditional wireless protocols [2].Currently there are several energy efficient communication models and protocols that are designed for specific applications and topologies.

LEACH (Low Energy Adaptive Clustering Hierarchy) is one of the most referenced protocols in the sensor networks area [3],[4],[5]. In LEACH and other routing protocols, when current cluster head changes due to self destruction or energy loss, increases the overhead, in turn leading to higher energy consumption. This is one of the worrying drawbacks. A possible solution which is proposed in this paper is the use of the bound time to reduce set-up communication overhead. During this bound time, sensor nodes receive advertisement messages and from this, node determines multi-route for transmission and consider only the message with the minimum number of hops and stand by cluster heads are elected.

The rest of the paper is organized as follows. In Section II, we review the related work. In Section III, describes Honey bee structure and modeling. The proposed method of BT –HBMA algorithm for cluster formation is described in Section IV. In Section V, We presented the Simulation results. Finally Section VI, concludes the paper.

II RELATED WORK

Hierarchical or Cluster –based routing, originally proposed in wire line networks, are well-known techniques with special advantages related to scalability and efficient communication. As such, the concept of hierarchical routing is also utilized to perform energy efficient routing in WSNs. In a hierarchical architecture, higher nodes can be used to process and send the information while low energy nodes can be used to perform the sensing in the proximity of the target. This means that creation of Clusters and assigning special tasks to cluster-heads can greatly contribute to overall system scalability, lifetime, and energy efficiency.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Hierarchical routing is an efficient way to lower energy consumption within a cluster and by performing data aggregation and fusion in order to decrease the number of transmitted messages to the base station.

Heinzelman[4] introduced a hierarchical clustering algorithm for sensor networks, called Low Energy Adaptive Clustering Hierarchy(LEACH). LEACH is a cluster based protocol, which includes distributed cluster formation. The operation of LEACH is split into two phases, the set-up phase and steady state phase. During the set-up phase, the clusters are created and cluster heads are elected. LEACH randomly selects a few sensor nodes as cluster-heads and broadcast an advertisement message to the entire network declaring that they are the new cluster heads. Every node receiving the advertisement decides to which cluster they wish to belong based on the signal strength of the received message. The sensor node sends a message to register with the cluster-head of their choice. Based on a TDMA approach, the cluster head assigns the time slot to registered node for sending the data.

During the steady-state phase, sensor nodes can start transmitting data to their respective cluster-head. The cluster head applies aggregation functions to compress the data before transmission to the sink. After a predetermined period of time spent on the steady-state phase, the network enters the set-up phase again and starts a new round of creating clusters.

Although LEACH is able to increase the network lifetime, there are still a number of issues about the assumptions used in this protocol. LEACH assumes that all nodes can transmit with enough power to reach the base station if needed and that each node has computational power to support different MAC protocols. Therefore, it is not applicable to networks deployed in larger regions. It also assumes that nodes always have data to send, and nodes located close to each other have correlated data. It is not obvious how the number of the predetermined cluster-head is going to be uniformly distributed through the network. Therefore, there is the possibility that the elected cluster head will be concentrated in one part of the network. Hence, some nodes will not have any cluster heads in their area.

Lindsey and Raghavendra[6],[7] proposed an enhancement over LEACH protocol. The protocol, called Power-Efficient Gathering in Sensor Information Systems (PEGASIS), is a near optimal chain –based protocol. It achieved the performance through the elimination of the overhead caused by dynamic cluster formation and through decreasing the number of transmissions and reception by using data aggregation. Although the clustering overhead is avoided, still requires dynamic topology adjustments.

This paper provides a protocol with the same underlying benefits as LEACH and PEGASIS and reduces the number of set-up messages required which in turn increases the network lifetime.

III HONEY-BEE COLONY STRUCTURE

A honey-bee colony typically consists of a single egg laying long-lived queen, anywhere from zero to several thousand drones (depending on the season) and usually 10,000 to 60,000 workers [8]. The colony can be founded in two different ways. In “independent founding” the colony starts with one or more reproductive females that construct the nest, lay the eggs, and feed the larvas. The first group of broods is reared alone until they take over the work of the colony. Subsequently, division of labor takes place and the queen specializes in egg laying and the workers in brood care . Another founding method is called “swarming” in which a new colony is founded by a single queen or more, along with a group of workers from the original colony.

A colony of bees is a large family of bees living in one bee-hive. A bee hive is like a big city with many “sections of the town”. The queen is the most important member of the hive because she is the one that keeps the hive going by producing new queen and worker bees. With the help of approximately 18 males (drones), the queen bee will mate with multiple drones one time in her life over several days. The sperm from each drone is planted inside a pouch in her body. She uses the stored sperms to fertilize the eggs. Whether a honeybee will become a queen, a drone, or a worker, depends on whether the queen fertilizes an egg. Since she is the only bee in the colony that has fully developed ovaries, the queen is the only bee that can fertilize the egg. Queens and workers come from fertilized eggs and drones from unfertilized eggs.

Only the queen bee is fed “royal jelly,” which is a milky-white colored jelly-like substance. “Nurse bees” secrete this nourishing food from their glands, and feed it to their queen. The diet of royal jelly makes the queen bee bigger than any other bees in the hive. A queen bee may live up to 5 or 6 years, whereas worker bees and drones never live more than 6 months. There are usually several hundred drones that live with the queen and worker bees. Mother nature has given the drones just one task which is to give the queen some sperm. After the mating process, the drones die. As the nights turn colder and winter knocks the door, the drones still in the hive are forced out of the hive by worker bees. It is a sad thing, but the hive will not have enough food if the drones stay.

Queens represent the main reproductive individuals which are specialized in eggs laying [9]. Drones are the fathers of the colony. They are haploid and act to amplify their mothers’ genome without altering their genetic composition, except through mutation. Workers are specialized in brood care and sometimes lay eggs. Broods arise either from fertilized or unfertilized eggs. The former represent potential queens or workers, whereas the latter represent prospective drones.

The mating process occurs during mating-flights far from the nest. A mating- flight starts with a dance where the drones follow the queen and mate with her in the air. In a typical mating-flight, each queen mates with seven to twenty drones. In each mating, sperm reaches the spermatheca and accumulates there to form the genetic pool of the colony. Each time a queen lays fertilized eggs, she retrieves at random a mixture of the sperms accumulated in the spermatheca to fertilize the egg. Insemination ends with the eventual death of the drone, and the queen receiving the “mating sign.” The queen mates multiple times but the drone inevitably only once. These features make bees-mating the most spectacular mating among insects.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A. Honey-bees modeling

The mating–flight may be considered as a set of transitions in a state-space (the environment) where the queen moves between the different states in some speed and mates with the drone encountered at each state probabilistically. At the start of the flight, the queen is initialized with some energy content and returns to her nest when her energy is within some threshold from zero or when her spermatheca is full.

In developing the algorithm, the functionality of workers is restricted to brood care and therefore, each worker may be represented as a heuristic which acts to improve and/or take care of a set of broods (i.e., as feeding the future queen with royal jelly). A drone mates with a queen probabilistically using an annealing function as[10]:

$$\text{Prob}(Q, D) = \exp\left[-\Delta(f) / s(t)\right] \quad (1)$$

where Prob (Q, D) is the probability of adding the sperm of drone D to the spermatheca of queen Q (that is, the probability of a successful mating); $\Delta(f)$ is the absolute difference between the fitness of D (i.e. $f(D)$) and the fitness of Q(i.e. $f(Q)$); and $S(t)$ is the speed of the queen at time t. It is apparent that this function acts as an annealing function, where the probability of mating is high when both the queen is still in the start of her mating–flight and therefore her speed is high, or when the fitness of the drone is as good as the queen's. After each transition in space, the queen's speed, $S(t)$, and energy, $E(t)$, decay using the following equations:

$$S(t+1) = a \times S(t) \quad (2)$$

$$E(t+1) = E(t) - \gamma \quad (3)$$

Where a *is* a factor $\in [0, 1]$ and γ is the amount of energy reduction after each transition. Thus, an Honey-Bees Mating Optimization (HBMO) algorithm may be constructed with the following five main stages [11]:

1) The algorithm starts with the mating–flight, where a queen (best solution) selects drones probabilistically to form the spermatheca (list of drones). A drone is then selected from the list at random for the creation of broods.
2) Creation of new broods (trial solutions) by crossoverring the drones' genotypes with the queen's.
3) Use of workers (heuristics) to conduct local search on broods (trial solutions).
4) Adaptation of workers' fitness based on the amount of improvement achieved on broods.
5) Replacement of weaker queens by fitter broods.

IV. PROPOSED METHOD

A. Bound Time (BT) Routing Protocol

The main objective of our BT approach is to minimize the set-up communication overhead, whenever current cluster head changes. These changes are due to cluster head failures or when its energy level approaches a certain threshold value. During the bound-time, sensor node receives single multi-purpose message and from this the node starts to determine the following 1) possible routes from the cluster head to sensor node 2) learns the minimum number of hops to reach the selected cluster head.3) Stand by nodes are chosen for next to current cluster head. Hence, this single multi-purpose advertisement message can be used for both reducing the set-up communication overhead and fault tolerant, thus makes our protocol more energy efficient.

The operation of the proposed routing protocol can be split into two phases: the role determination phase and the data transfer phase

B. Role Determination Phase

During this phase, cluster heads are selected and clusters are formed. At the start up, base station randomly selects some desired percentage of nodes as cluster heads and broadcasts selected information to the network. On receiving the broadcasted information, each node checks its status whether it has been selected as cluster head or not. If yes, it starts a new cluster formation by broadcasting an advertisement message. Otherwise, it forwards the message to its neighbors. Every cluster head creates an advertisement message which has the number of hops count to zero and broadcast it to its neighbors. If a node already belongs to another cluster for which the number of hops to reach the current belonging cluster is less than newly received broad cast then it ignores the received message.

The bound time of a node starts when it accepts an advertisement message. When the bound-time is still valid, the node caches the received message and waits for other possible advertisement. In this way, it collects all possible alternative paths to chosen cluster head. All the sensor nodes consider the message with minimum number of hops count (shortest route) as the best route. When route fails, an alternate route can be immediately used without delays or degradation of QoS.When the bound-time reaches zero, a route is established with shortest route and increases the number of hops count by one in the retained message and broadcasts it to its nearby nodes.

After bound time expires, all sensor nodes who receive the advertisements message are candidate for stand by node to their respective cluster head. All the sensor nodes who expressed their willingness are collected in the stand by node list and stored in the cluster head. This stand by node list is used as input to our proposed Honey mating algorithm which is discussed in next section. In the meantime, data transfer phase is started for conducting data transfer in the network.

The current energy of the current cluster head is polled in every round time. When the current cluster head energy is depleted to near specified threshold energy level, our proposed Honey Bee Mating algorithm is triggered to find the best stand by node for current cluster head from the stand by node list. When current cluster head about to dead completely, the best stand by node selected using our approach replaces the current cluster head. This newly elected cluster head can withstand for many rounds and there by reduces the number of re-election.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Honey-bee mating behavior discussed in section III is equivalently mapped to our proposed BT-Honey Bee mating algorithm in electing the stand by cluster head as shown in figure 1 & figure 2.

..

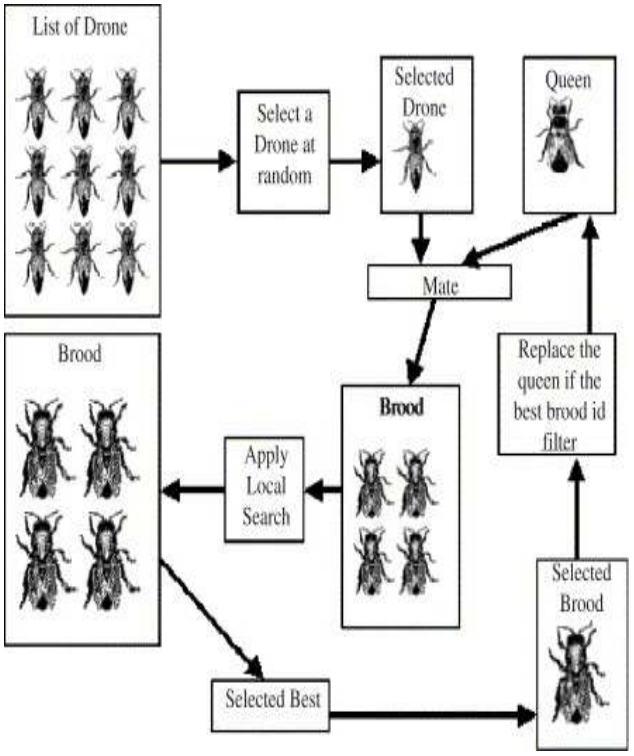


Figure 1. Real Honey Bee Mating Approach

C. Honey Bee Mating Algorithm.

1. After the Bound time expires, Cluster head collects the nodes who expressed willingness to act as stand by node and creates the node list called Trial Stand by node List.

2. Before the start of next step, current energy of nodes in the stand by node list are examined and nodes having less energy are discarded from the stand by node list.

3. From the Standby Node List, Best Node and Next Best Node List are created using the objective function. The Best node List and Next best node list is equivalent to queen and drones respectively. Our objective function is based on the remaining energy of stand by nodes. Obviously, the higher the enrgy,the higher the probability that the node will become cluster head.

The remaining energy is calculated as

$$E_r = \frac{E_c}{E_i} \qquad (4)$$

Where E_c is the current energy and E_i is the initial energy.

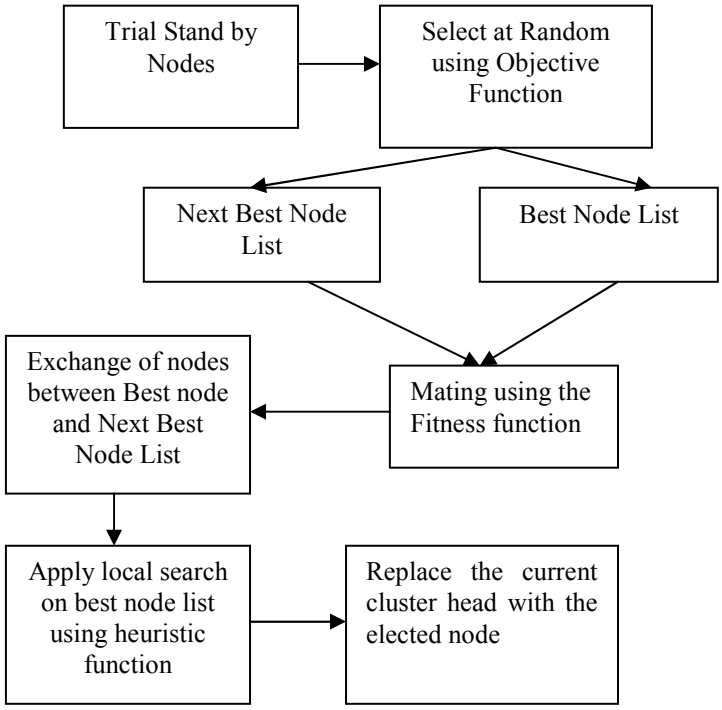


Figure 2. Honey Bee Mating Approach

4. Mating is done through exchange (crossover) between Best Node and Next Best Node List based on the fitness function which depends on the closeness to the sink node (if selected).The new best node list are broods. Our fitness function is based on the closeness of stand by node from other nodes if it is selected as new cluster head. In cluster, calculates the distance of each node to its cluster-head (if selected) based on number of hops. The lesser the distance the higher the probability that the node will become cluster head. The number of hops between i and j node is calculated as follows

No_of_hops (i, j) =

$$\xrightarrow[A->1..n]{Min} \left[no\_of\_hops(i, I_{A1}) + \sum_{k=1}^{j-1} no\_of\_hops(I_{AK}, I_{A(k+1)}) \right] \qquad (5)$$

$A->1..n$ = Number of Alternative path between i and j node.

I_{A1} = 1st Intermediate node in the A th Alternative path to node j

K varied from 1st node to j-1 (previous node in reaching jth node) node.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

5. Perform a local search on the Best Node List by applying the heuristic function which depends on the distance between cluster head [if selected] and base station. The lesser the distance, less power is needed to transmit and receive the data.. Such nodes are having higher probability to become the cluster head. This local search gives the best solution.

6. In case of re-election, current cluster-head is replaced by new cluster -head which is selected in step 5.

When the cluster round time is over or energy level approaches a threshold, the current cluster head hands the main role to new elected cluster head . With a single flooding to cluster members, the new cluster head continues its main role without the need for further communication.

D. Data Transfer Phase

The second phase is called data transfer or steady state phase. In this phase, data transfer is started as soon as bound time expires. This phase is identical to steady state phase proposed in LEACH[5].The steady state is broken into frames where nodes send their data to the cluster-head at most once per frame during their allocated transmission slot which scheduled by TDMA. Once the cluster- head receives all the data, it performs data aggregation and forwards the final data to the base station.

V SIMULATION RESULT

Our proposed algorithm is implemented and completely studies using the simulation tool called SenSor Plus. This is an in- house sensor network simulator. SenSor [12] is a realistic and scalable Python based simulator that provides a workbench for prototyping algorithms for WSN. It consists of a fixed API, with customizable internals. Each simulated sensor node runs in its own thread and communicates using the same protocols as its physical counterpart would be. This enables experimentation with different algorithms for managing the network topology, simulating fault management strategies and so on, within the same simulation. SenSor Plus is an extension of Sensor with an added interface between the simulation environment and different hardware platforms, for example the *Gumstix* [13] platform. SenSor Plus bridges between Sensor and the Gumstix to allow applications implemented within the simulator to be ported directly on to the hardware. Sensors are modeled using a pool of

concurrent, communicating threads. Individual sensors are able to:

1) Gather and process data from a model environment
2) Locate and communicate with their nearest neighbours
3) Determine whether they are operating correctly and act accordingly to alter the network topology in case of faulty nodes being detected.

Separate interfaces gather information from the network and display it on the graph pane or the chart pane, where individual data can be plotted during the simulation. This partitioning allows us to experiment with different ways of processing individual node data into information

Using the SenSor Plus framework, we implemented the proposed algorithm. For our simulation, we gave all the nodes an initial supply of energy and ran the protocol until it converged. For our Experiments, we created a 100-node network, where the nodes are scattered randomly on 600×600 grid, such that no two nodes share the same location. In our simulation, we considered the initial simulation parameter and its values as shown in Table I.

TABLE I SIMULATION PARAMETERS

Parameter	Value	Motivation
Initial Energy	1 J	Standard energy value used for batteries in most sensor nodes.
Broadcast size packet	11 bit	Assume that the sensor is broadcasting an ID of 11 bits.
Routing Size Packet	11 bit	Assume that the sensor is routing packets of size 11 bits

Our proposed BT-HBMA methods outperforms the other methods such as LEACH, PEGASIS in achieving reduced number of re election and maintains high energy nodes between the rounds. According to Table II, it is clear that our honey bee mating approach selects the best stand by node which can withstand for more rounds compare to other methods.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

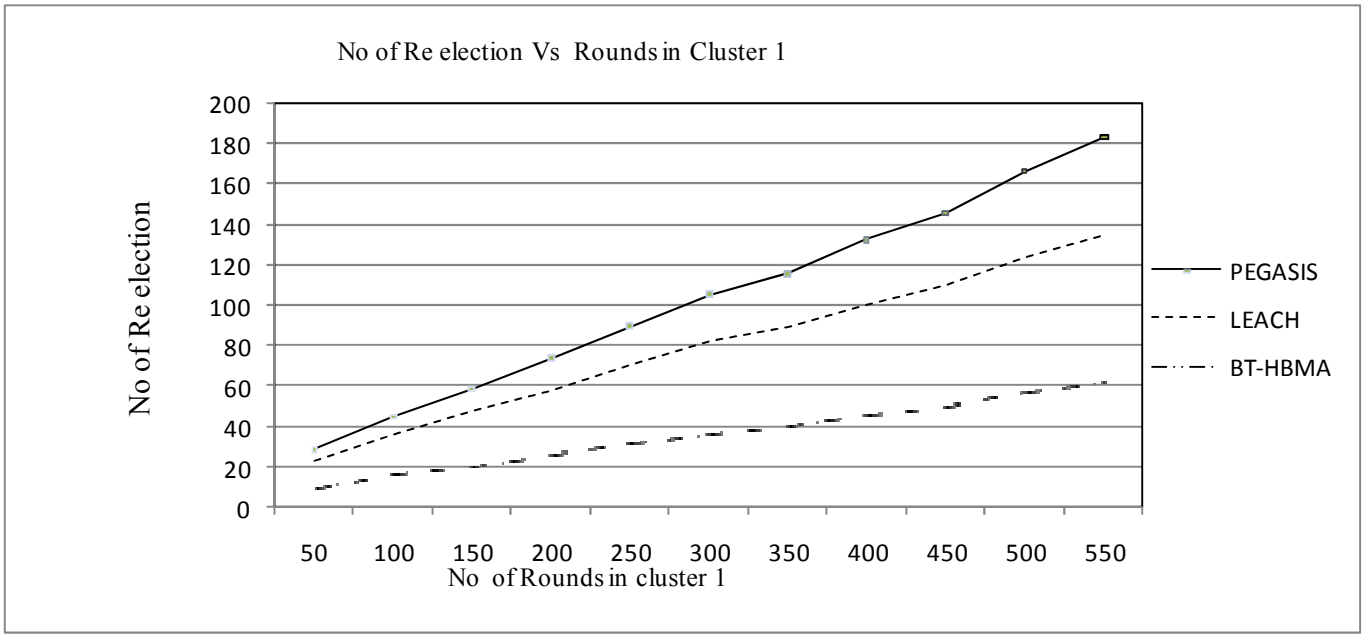


Figure3 No of Re election Vs No of Rounds in cluster 1

TABLE II NUMBER OF REELECTION BETWEEN THE ROUNDS IN CLUSTER1

Cluster 1 (Number of Rounds)	Number of Re-election in Cumulative Manner		
	LEACH	PEGASIS	BT-HBMA
150	21	27	11
250	32`	39	19
350	41	48	27
450	51	59	36
550	63	72	49

By this approach, we can reduce the total number of re election that likely to happen in the network if other methods are followed. This also implicitly achieves excellent data transfer phase. The Figure 3 depicts that our proposed method outperforms the other methods in reducing the number of re-election.

According to table III, the average energy of nodes in the cluster 1 is fair in our approach compared to other methods. The figure 4 shows that average energy of nodes in cluster 1 in terms of Joules between the numbers of rounds

TABLE III AVERAGE ENERGY OF NODES IN CLUSTER1 BETWEEN THE ROUNDS

Cluster 1 (Number of Rounds)	Average energy of nodes in cluster1 (Joules)		
	LEACH	PEGASIS	BT-HBMA
150	0.97	0.95	0.982
250	0.95	0.93	0.962
350	0.84	0.81	0.895
450	0.752	0.69	0.842
550	0.593	0.43	0.712

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

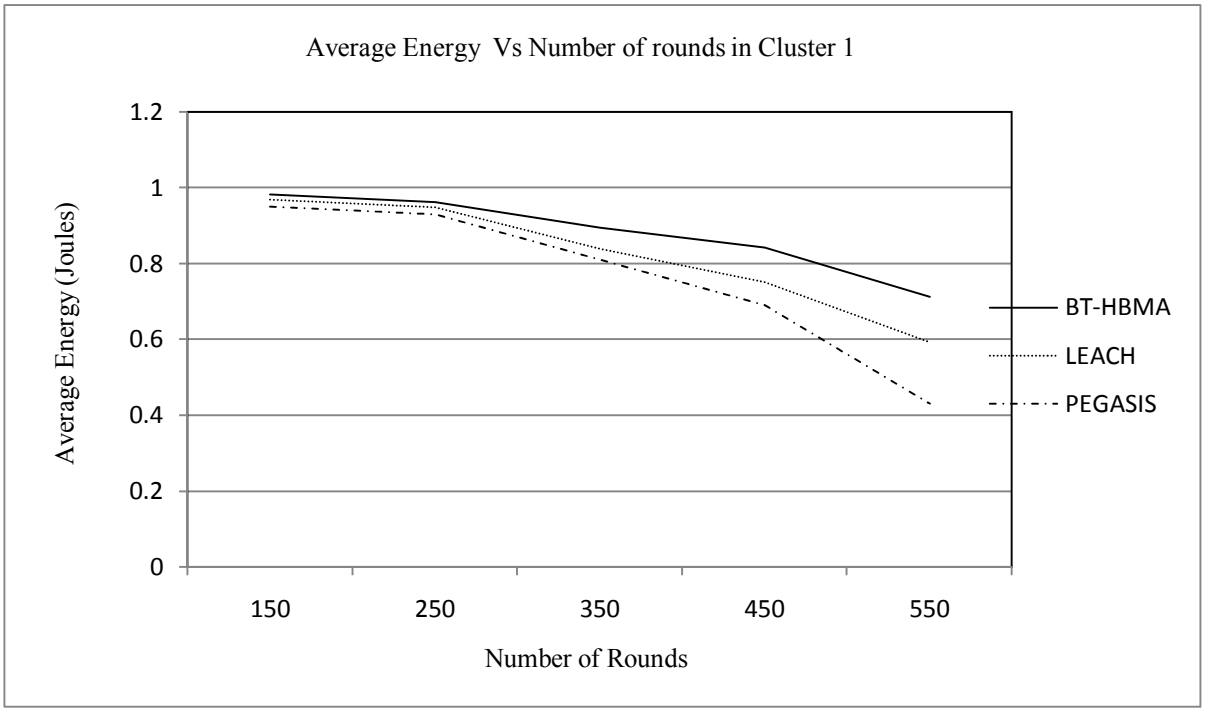


Figure 4 Average Energy of nodes in cluster 1 Vs No of Rounds in cluster 1

Our proposed BT-HBMA method, each cluster head and its respective members are having higher average energy even after the 550 rounds. The other methods decay energy because of the more number of re election.

VI. Conclusions

This paper provides a better solution to reduce the number of re-elections by selecting the best stand by node in advance for current cluster head. Our proposed BT-HBMA is equivalently mimics the Honey Bee Mating Behaviour of Real Honey Bees. We have applied the Honey mating behaviour for the first time in the stand by node selection of Wireless sensor network. This Honey bee mating inspired approach is effectively used to choose the best stand by node from multiple available solutions. Our proposed BT-HBMA method is accurately finds the best stand by node from available Trial stand by node list. The stand by node choosen as best solution in our approach withstands for many rounds and improves over all data transfer rate as well as maintains fair average energy in the Wireless Sensor Network. Our simulation results shows that there is reduction in number of re election conducted in the network compared to the existing methods like LEACH and PEGASIS. In future, we would like to apply our BT-HBMA to different Wireless Sensor Network Layouts for improving the scalability factor.

References

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," Computer Networks, vol. 38, no. 4, 393– 422, 2002.

[2] D. Estrin, D. Culler, K. Pister, and G. Sukhatme, "Connecting the physical world with pervasive networks," IEEE Pervasive Computing, 59–69, January-March 2002.

[3] W.B Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," In the Proceedings of the 33rd Hawaii International Conference on System Sciences (HICSS '00), January 2000.

[4] A.Manjeshwar and D.P. Agarwal, "TEEN: a routing protocol for enhanced efficiency in wireless sensor networks," In the 1st International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing, April 2001.

[5] M. Ye, C. Li, G. Chen, and J. Wu. Eecs: An energy efficient c lustering scheme in wireless sensor networks. In Performance, Computing, and Communications Conference, 2005. IPCCC 2005. 24th IEEE Interna tional, pages 535– 540, April 2005.

[6] S. Lindsey and C. Raghavendra, "PEGASIS: Power-Efficient Gathering in Sensor Information Systems," IEEE Aerospace Conference Proceedings, 2002, vol. 3, no. 9-16, 1125-1130.

[7] M. Hammoudeh. Robust and energy efficient routing in wire less sensor networks. Master's thesis, University of Leicester, 2006.

[8] R.F.A. Moritz and E.E. Southwick, Bees as Superorganisms, Springer, Berlin, Germany (1992).

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[9] H.H. Laidlaw and R.E. Page, Mating designs. In: T.E. Rinderer, Editor, Bee Genetics and Breeding, Academic Press Inc., New York, NY (1986), pp. 323–341.

[10] H.A. Abbass, Marriage in honey-bee optimization (MBO): a haplometrosis polygynous swarming approach, in: The Congress on Evolutionary Computation (CEC2001), Seoul, Korea, May 2001, 2001, pp. 207–214.

[11] O. Bozorg Haddad, A. Afshar, M.A. Mariño, Honey bees mating optimization algorithm (HBMO); a new heuristic approach for engineering optimization, in: Proceeding of the First International Conference on Modeling, Simulation and Applied Optimization (ICMSA0/05), Sharjah, UAE, 1–3 February 2005.

[12] S. Mount, R. Newman, and E. Gaura. A simulation tool for system Conference and Trade Show, 3:423, May 2005.

[13] Gumstix.com. Gumstix way small computing, 2007. [Online; accessed 26-March-2007].

AUTHORS PROFILE

J.Senthilkumar received his ME Applied Electronics with Distiction from Anna University, Chennai in 2004. He is working as Assistant professor in IT Department of Sona College of Technology and pursuing PhD degree in Anna University, Chennai. He is active in the research area of Wireless Sensor Networks, Mobile computing and Data mining.

Dr.R.Lakshmipathi received his BE from College of Engineering, Anna University, India, in 1971, the ME and PhD in Electrical Engineering from College of Engineering, Anna University, India in 1973 and Indian Institute of Technology (IIT), Chennai in 1979, respectively. He has 36 years of teaching experience at UG degree level out of which 10 years in PG degree level. He worked as principal in Govt. College of Engineering and held the various prestigious posts like Dean, Regional Research Director , Chairman for board of BE exams, Member of Academic auditing committee, AICTE, University and State Govt. expert committee member. He is currently a professor of Electrical Engineering, St. Peters Engineering College (Deemed University), Tamilnadu. His research interest includes Electrical power semi conductor drives, Signal processing and Web Mining.

V.Mohanraj received his ME Computer science and Engineering from Anna University, Chennai in 2004. He is currently working as Assistant professor in IT department of Sona College of Technology. He is pursuing PhD degree in Anna University, Chennai. His research area includes web mining, database and intelligent system.

Y.Suresh received his ME Applied Electronics from Anna University, Chennai in 2004. He is working as Assistant professor in IT Department of Sona College of Technology and pursuing PhD degree in Anna University, Chennai. He is active in the research area of Computer Netoworks, Data mining, control system and Mobile computing.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Multi-Label Image Segmentation for Medical Applications Based on Graph- Cuts

R.Rajeswari,
Research Scholar, Department of Information Technology
MIT campus, Anna University
Chennai-44,India
rajimaniphd@gmail.com

P. Anandhakumar,
Asst. Professor, Department of Information Technology
MIT Campus, Anna University
Chennai-44, India
anandh@annuniv.edu

***Abstract* : Mathematical morphology is very attractive for automatic image segmentation because it efficiently deals with geometrical descriptions such as size, area, shape, or connectivity that can be considered as segmentation-oriented features. This paper presents a new approach for some theoretical and practical improvements of image segmentation based on graph cuts. The method is based on the use of the region adjacency graph produced by the watershed transform from mathematical morphology. Marker extraction identifies the presence of homogeneous regions. The combination of morphological and graph cuts segmentation permits us to speed up and define new classes of energy functions that can be minimized using graph cuts. The use of region graphs gives promising results and can potentially become a leading method for interactive medical image segmentation.**

***Keywords*- Image segmentation, marker extraction, morphology, Graphs, Watershed Transform.**

I. INTRODUCTION

MAGNETIC resonance imaging (MRI) provides detailed images of living tissues, and is used for both brain and body human studies. Data obtained from MR images is used for detecting tissue deformities such as cancers and injuries. The key to any automatic method is that it must be robust, so that it produces reliable results on every image acquired from any MR scanner using different relaxation times, slice thicknesses and fields of view. More recently, computer-assisted methods have been used for specific tasks such as extraction of MS lesions from MRI brain scans [1], [2].

A key problem in medical imaging is automatically segmenting an image into its constituent heterogeneous processes. Automatic segmentation has the potential to positively impact clinical medicine by freeing physicians from the burden of manual labeling and by providing robust, quantitative measurements to aid in diagnosis and disease modeling. One such problem in clinical medicine is the automatic segmentation and quantification of brain tumors. The model-aware affinities integrated into the multilevel segmentation by weighted aggregation algorithm, and apply the technique to the task of detecting and segmenting brain tumor and edema in multichannel magnetic resonance (MR) volumes.[3].

The flowchart of proposed system shown in Fig.1.The tumors segmentation is obtained through the computation of a minimal graph cut of the region adjacency graph obtained on the watershed segmentation of the image. The user has to interactively specify the location of the tumors as well as a marker to specify the background. A post-processing step is also proposed to smooth the segmentation result. This post-processing step consists in a morphological opening of the segmentation.

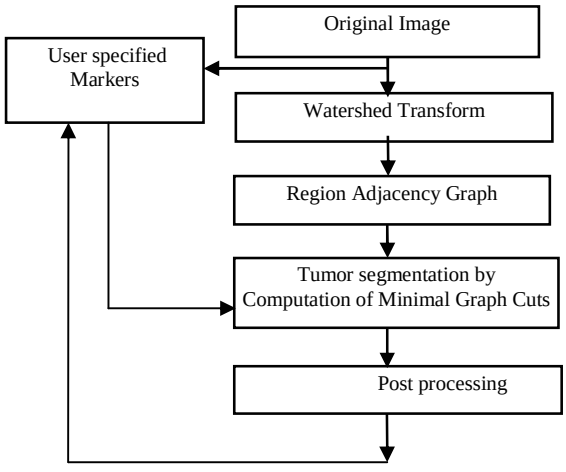


Fig 1.Flow Chart of proposed method

II. RELATED WORK

There are lot of work available related to the proposed work, in which some of them are significant. It includes

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

automatic and semiautomatic for segmentation of brain into different tissues, including MS lesions. These approaches include a variety of methods such as statistical, fuzzy, neural networks, and fuzzy neural networks.

Kamber et al. [4] developed a brain tissue model for segmentation of MR images of patients with MS disease. The offered model was three-dimensional, voxel based, which provided *a priori* probabilities for white matter, gray matter and CSF.

The survey by Clarke *et al*. of segmentation methods for MR images [5] describes many useful image processing techniques and discusses the important question of validation. The various image processing techniques used for segmenting the brain can be divided into several groups: those required to perform a crude threshold-based extraction of the brain, followed by refinement of brain contours; statistical methods for brain segmentation, and region growing methods.

The work by M. Kass, A. Witkins & D. Terzopoulos [6] in 1988 is the basis for the active contour model, the Snake algorithm. The work by Mathews Jacob, Thierry Blu and Michael unser [7] presents different solutions for improving spline based snakes, and demonstrates the minimum curvature interpolation property used as an argument to get rid of the explicit smoothness constraint.

The work of El Naqa et al [8] proposes a variational methods based on multivalued level set deformable models for simultaneous 2D or 3D segmentation of multimodality images or multiple image sets from the same modality.

A multidimensional segmentation and filtering methodology for accurate blood flow velocity field reconstruction from phase-contrast magnetic resonance imaging (PC MRI) developed and validated by Kartik S. Sundareswaran, David H. Frakes et.al[9].

Yuri Boykov, Olga Veksler and Ramin Zabih [10] proposes an expansion algorithm that finds a labeling within a known factor of the global minimum, while swap algorithm handles more general energy functions. Both algorithms allow important cases of discontinuity preserving energies for image restoration, stereo and motion.

Saracoglu *et al*. [11] modeled the problem using a method consisting of three steps: image tessellation, clustering, and classification. The image was tessellated into regions with similar properties using a region growing approach (tessellation step).Based on the “average” color information of the regions, clustering is performed.

A novel method for segmentation and classification of M-FISH chromosome images is presented by Petros S. Karvelis, Alexandros T. Tzallas et al.[12]. The segmentation is based on the multichannel watershed transform in order to define regions of similar spatial and spectral characteristics. Then, a Bayes classifier task-specific on region classification, is applied. The combination of the multichannel segmentation and the region-based classification is found to improve the overall classification accuracy compared to pixel-by-pixel approaches**.**

A very powerful segmentation method that has been widely used in image segmentation problems is the watershed transform (WT) [13]–[19]. Since its original application on grayscale images [13], [14], a very attractive computational form has been derived [15] and extended to color images [16]. An extensive review of the watershed algorithms can be found in [19].The watershed transform presents some advantages over other developed segmentation methods.

1) The watershed lines form closed and connected regions, where edge based techniques usually define disconnected boundaries that need postprocessing to produce closed regions.

2) The watershed lines always correspond to obvious contours of objects which appear in the image.

The main problem of over-segmentation, can be usually overcome by the use of preprocessing or postprocessing, producing a segmentation that better reflects the arrangement of objects within the image.

III. WATERSHED BASED IMAGE TRANSFORM

The proposed system starts with the original image and watershed transform of the image's gradient modulus is used for segmentation purpose.

The validation is done comparing our segmentation with the physician’s delineation. Fig 1 shows the flowchart of the proposed method.

A. Watershed Based on Dissimilarity Measures

The watershed transform definitions can be slightly modified in such a way that it produces a segmentation according to different path based criteria. For image segmentation purposes the watershed transform is computed on the image's gradient modulus, usually the morphological gradient. The estimation of the gradient has thus also an importance for the quality of the segmentation. A first extension of the classical watershed transform of the gradient image is based on the local dissimilarity between neighbor pixels. This approach, called watershed by dissimilarity was originally proposed by Pard as[20] and Lotufo et.al in [21].

Given a pixel adjacency graph G = (V;E;W), we consider the following edges weights mapping

$$e_{i,j} \in E, w_{i,j} = (1/d(i,j).|pi—pj)|+1)$$

where i and j are two nodes of the graph, pi and pj are the grey level values of neighbor pixels of the image and d(i, j) is the distance between the two pixels.

These edges weights represent the local estimation of the image's gradient modulus. Instead of computing the morphological gradient of the image at each pixel, the gradient is here estimated "between" the pixels. This weight map especially provides a better detection of thin contrasted objects.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

This weight map has the main advantage to achieve a better detection of small details compared to the classical morphological gradient.

IV. GRAPHS IN IMAGE SEGMENTATION

This section presents the graphs commonly encountered in imaging applications. A first level is the so-called "pixel adjacency graph". In a second level, an unsupervised low level segmentation of the image (i.e. a segmentation that provides much more regions than objects in the image) is used to build a "region adjacency graph".

A. *Pixel Adjacency Graphs*

In this case, the set of nodes of the graph is the set of pixels of the image, and the edges link neighboring pixels. For each image and an adjacency system is needed to build the corresponding pixel adjacency graph. Different common adjacency systems are illustrated in figure.

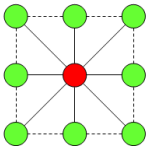


Figure 1. 4 neighborhood adjacency system (V4).

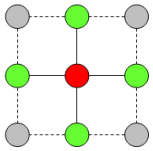


Figure 2. 8 neighborhood adjacency system (V8).

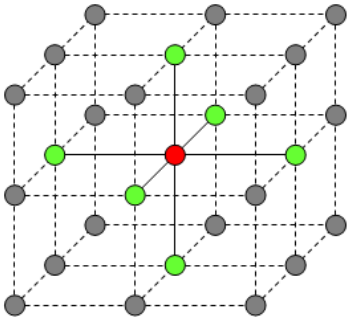


Figure 3. 6 neighborhood adjacency system for 3D

B. *Region Adjacency Graph*

A region adjacency graph can be obtained from low-level segmentation of the image by numerous methods. It can be obtained from a first unsupervised clustering of the image, for instance one can apply the watershed transform [22, 23], flat zones labeling [24, 25], or k-means clustering [26] to obtain such low-level segmentations.

The minimal cut problem on region adjacency graph is used instead of pixel adjacency graph since it is advantageously used to speed up and extend the presented techniques.

V. IMPLEMENTATION AND RESULTS

The region adjacency graph is obtained from the watershed transform computed from all minima of the morphological gradient of the original MRI image using Meyer's algorithm. [27]. Then a region adjacency graph is extracted and used for the next optimization steps. The user specified markers are used to compute a minimal graph cut separating the markers specifying the myocardium and the external tissues. The markers that specify the myocardium denoted as the set of regions Mm. The markers specifying the tissues surrounding the myocardium are denoted by $M_{ext.}$

$F(r_i , r_j)$ is defined as the set of edges of the pixel graph connecting two regions r_i and r_j of the low-level watershed segmentation.

$$F(r_i , r_j)=\{e_{m,n} \in E \mid m \in r_i, n \in r_j\} \quad (1)$$

The strictly positive and decreasing function g used as an edge indicator for the a minimal surface is

$$g(\|I(p)\|)= (1/(1+\|I(p)\|)^k \quad (2)$$

where the k parameter was set to 2.

The edges weights of the region adjacency graph $G_{R=} (V_R, E_R, W_R)$ are then set such that the weight of a graph cut equals the energy function of a surface defined as

$$^{w} r_{i,} , r_j = \Sigma (g(\| I(p)\|)) \quad (3)$$

The myocardium boundaries are finally extracted by computing a minimal graph cut of the region adjacency graph with weights given by (3). The minimal cut is obtained on the region adjacency graph with two additional nodes s and t, respectively connected to the markers of the myocardium and the markers of the external tissues. The edges weights of the graph are summarized in Table I.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

TABLE I. EDGE WEIGHTS OF GRAPH

Edge	Weight	for
$^{w}s, r_i$	$+\infty$	$r_i \in M_m$
$^{w} r_i, t$	$+\infty$	$r_j \in M_{ext}$
$^{w} r_i, r_j$	Σ $(g(\|\| I(p)\|\|))$ $(e_{m,n} \in F(r_i, r_j))$	$r_i \in V_R, r_j \in N_r$

A. Dataset

The imaging modality in this application is an isotropic 3D MRI. This imaging technology provides 50 to 120 slices covering the whole heart. The whole myocardium is not always visible on 3D MRI datasets, especially along the right chamber, because the myocardium is too thin in this area. Thus the left part of the myocardium which is outlined in green in Fig is segmented in our proposed method. The other parts of the myocardium have been obtained with manual segmentations.

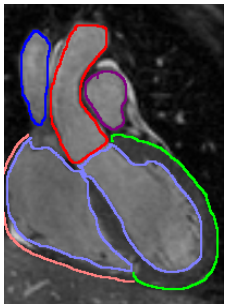


Figure 4a. A slice of a 3D Heart MRI

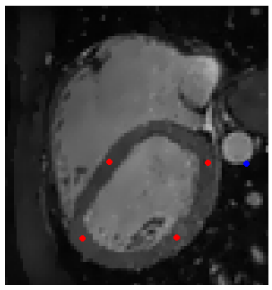


Figure 4b. MRI heart image superposed with markers

Figure 4c. Minimal surface computed by proposed method

A slice of a 3D Heart MRI superposed with a hand made segmentation is shown in figure 4a. The segmentation highlights the aorta (in red), the superior vena cave (in dark blue), the pulmonary artery, the right and left chamber (light blue) and the heart muscle (green and light red).

C. Segmentation

For each image, a reference segmentation done by an experienced radiologist is available.

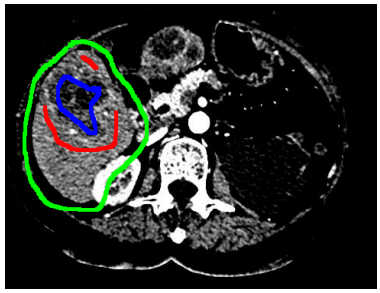


Figure 5a. Liver tumors segmentation. (a) User specfied markers.

Figure 5b. Results of proposed segmentation strategy

In order to evaluate the accuracy of the developed method, four scores are computed from five different evaluation measures.

1) Volumetric overlap: This score is computed as the number of voxels in the intersection of segmentation and reference, divided by the number of voxels in the union of segmentation and reference. This value is equal to 1 for a perfect segmentation and is equal to 0 as the lowest possible

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

value, when there is no overlap at all between the segmentation and the reference.

2) Relative absolute volume difference: The total volume of the segmentation is divided by the total volume of the reference. From this number 1 is subtracted, the absolute value is taken and the result is multiplied by 100.

3) Average symmetric absolute surface distance: These are defined as those voxels in the object that have at least one neighbor that does not belong to the object. For each voxel in these sets, the closest voxel in the other set is determined. All these distances are stored, for boundary voxels from both reference and segmentation. Taking the average of all these distances symmetric absolute surface distance is determined.

4) Symmetric RMS surface distance: The squared distances between the two sets of border voxels is taken and the root is extracted and gives the symmetric RMS surface distance.

D. Comparison of Results:

1) Training: The evaluation scores of our method calculated on a set of MRI images presenting different tumors with known handmade segmentations. The evaluation scores compare our results with the radiologists segmentations. The important point is that these results have been obtained with the knowledge of the handmade segmentations. The results have been obtained such that the similarity between the two segmentations is visually satisfactory.

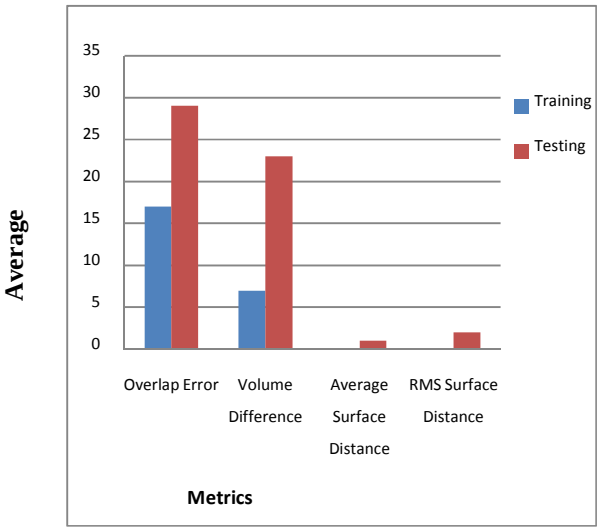


Figure 1. Comparison between Training and Testing Results

VI. CONCLUSION

The developed tools, based on low-level segmentation and graph based methods, have been successfully used for various medical applications including tumors delineation and model creation for surgery planning and simulation. The challenging problems presented by medical applications forced us to imagine innovative methods that can tackle the problems linked with clinical applications. The method presented have all the characteristics to be used in real life applications: robustness, speed and precision.

VII FUTURE ENHANCEMENT

The minimal cuts can be extended to provide constrained segmentation models. These new constrained problems are solved in a linear programming framework. The major limitation of linear programming is the current inefficiency of generic solvers. An efficient linear program solver that can handle graphs representing large images can be developed.

REFERENCES

[1] Johnston, M. S. Atkins, B. Mackiewich, and M. Anderson, "Segmentation of multiple sclerosis lesions in intensity corrected multispectral MRI," IEEE Trans. Med. Imag., vol. 15. pp. 154–169, Apr. 1996.

[2] A. P. Zijdenbos, B. M. Dawant, R. A. Margolin, and A. C. Palmer,"Morphometric analysis of white matter lesions in MR images: Method and validation," IEEE Trans. Med. Imag., vol. 13, p. 4, pp. 716–724,Dec. 1994.

[3] Jason J. Corso, Eitan Sharon, Shishir Dube, Suzie El-Saden, Usha Sinha, and Alan Yuille "Efficient Multilevel Brain Tumor Segmentation With Integrated Bayesian Model Classification,"IEEE Trans. Med. Imag.,, vol. 27, No. 5, MAY , pp.629-640, 2000.

[4] M. Kamber, R. Shinghal, D.L. Collins, G.S. Francis, A.C. Evans, "Model-based 3D segmentation of multiple sclerosis lesions in magnetic resonance brain images,"IEEE Trans. Med. Imag. 14 (1995) 442–453.

[5] L. P. Clarke, R. P. Velthuizen, M. A. Camacho, J. J. Heine, M.Vaidyanathan, L. O. Hall, R. W. Thatcher, and M. L. Silbiger, "MRI segmentation: Methods and applications," Magn. Reson. Imag., vol. 13,no. 3, pp. 343–368, 1995.

[6] M. Kass, A. Witkins and D. Rerzopoulos, "Snakes: Active Contour Models", International Journal of Computer Vision, vol.1, p. 321-331, 1998.

[7] Mathews Jacob, Thierry Blu and Michael Uhnser, "A unifying approach and interface on Spline-Based Snakes,",Proceedings of SPIE, vol.4332, 2001, pp 340-347.

[8] Issam El Naqa, Deshan Yang, Aditya Apte, Divya Khullar, Sasa Mutic, Jie Zheng,Jeffrey D. Bradley, Perry Grigsby, and Joseph O. Deasy, "Concurrent multimodality image segmentation by active contours for radiotherapy treatment planning ,"Medical Physics**,** Vol. 34, No. 12, December 2007.

[9] Kartik S. Sundareswaran, David H. Frakes, Mark A. Fogel,Dennis D. Soerensen, John N. Oshinski, and Ajit P. Yoganathan, "Optimum Fuzzy Filters for Phase-Contrast Magnetic Resonance Imaging Segmentation", Journal of Magnetic Resonance Imaging 29:155–165 (2009).

[10] K. Saracoglu, J. Brown, L. Kearney, S. Uhrig, J. Azofeifa, C. Fauth, M.Speicher, and R. Eils, "New concepts to improve resolution and sensitivity of molecular cytogenetic diagnostics by multicolor fluorescence in situ hybridization," *Cytometry*, vol. 44, no. 1, pp. 7–15, May 2001.

[11] Petros S. Karvelis, , Alexandros T. Tzallas, Dimitrios I. Fotiadis, and Ioannis Georgiou, "A Multichannel Watershed-Based Segmentation Method for Multispectral Chromosome Classification," *IEEE Trans. Med. Imag.*, vol. 27, no. 5, pp.697–708, May 2008.

[12] S. Beucher and C. Lantuejoul, "Use of watersheds in contour detection,"in Proc. Int. Workshop Image Process., Real-Time Edge Motion *Detection/Estimation*, Rennes, France, 1979, pp. 17–21.

[13] S. Beucher, "The watershed transformation applied to image segmentation,"*Scan. Micros. Intern.*, vol. 6, pp. 299–314, 1992.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[14] L. Vincent and P. Soille, "Watershed in digital spaces: An efficient algorithm based on immersion simulations," IEEE Trans. Pattern Anal.Mach. Intell., vol. 13, no. 6, pp. 583–598, Jun. 1991.

[15] F. Meyer, "Color image segmentation," in Proc. Int. Conf. Image Process. Appl., 1992, pp. 303–306.

[16] V. Grau, A. U. Mewes, M. Alcaniz, R. Kikinis, and S. K. Warfield,"Improved watershed transform for medical image segmentation using prior information," IEEE Trans. Med. Imag., vol. 23, no. 4, pp.447–458, Apr. 2004.

[17] K. Haris, S. Efstratiadis, N. Maglaveras, C. Pappas, J. Gourassas, and G. Louridas, "Model-based morphological segmentation and labeling of coronary angiograms," IEEE Trans. Med. Imag., vol. 18, no. 10, pp.1003–1015, Oct. 1999.

[18] J. Roerdink and A. Meijster, "The watershed transform: Definitions,algorithms and parallelization strategies," Fundamenta Informaticae,vol. 41, pp. 187–228, 2001.

[19] M. Pardas, Segmentacion morfologica de secuencias de imagenes: Aplicacion a la codificacion, PhD thesis, Universitat Politcnica de Catalunya,). Barcelona, Spain. (1995).

[20] Lotufo and A. Falcao, The ordered queue and the optimality of the watershed approaches, Proceedings of the International Symposium on Mathematical Morphology 2000, Palo Alto, CA, Jun 2000. (2000), 341-350.

[21] N. Biggs, E. Lloyd, and R. Wilson, Graph theory, 1736-1936, Oxford University Press, 1986.

AUTHORS PROFILE

Dr.P.AnandhaKumar is Currently working as AssistantProfessor in the Departm-ent of Information Technology at Madras Institute of Technology, Anna University, Chennai,India. He has published several papers in international, national journals and conferences. His areas of interests includes content-based image indexing techniques and frameworks, image processing and analysis, video analysis ,fuzzy logic, pattern recognition, knowledge management and semantic analysis

R.Rajeswari is currently doing Ph.D in the Department of Information Technology, MIT Campus,AnnaUniversity Chennai,India

[22] ___, "The morphological approach to segmentation: the watershed transformation, Mathematical Morphology in Image Processing" (E.R. Dougherty, ed.), 1993, pp. 433-481.

[23] P. Salembier and J. Serra, Flat zones filtering, connected operators and filters by reconstruction, IEEE Transactions on Image Processing 3 (1995), no. 8, 1153-1160.

[24] J. Crespo, R. Schafer, J. Serra, C. Gratin, and F. Meyer, The flat zone approach: a general low-level region merging segmentation method, Signal Processing 62 (1997), 37-60.

[25] J. McQueen, Some methods for classification and analysis of multivariate observations, In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1 (1967),281-297.

[26] F. Meyer and S. Beucher, Morphological segmentation, Journal of Visual Communication and Image Representation 1 (1990), no. 1, 21-46.

Hybrid Model of Texture Classification using 2D Discrete Wavelet Transform and Probablistic Neural Network

Reem Abd El-Salam El-Deeb
Department of Computer Science,
Faculty of Computer and
Information Sciences,
Mansoura University, Egypt,
P.O.BOX:35516
Reemm_db@yahoo.com

Taher. Hamza
Department of Computer Science,
Faculty of Computer and
Information Sciences,
Mansoura University, Egypt,
P.O.BOX:35516
Taher_hamza@yahoo.com

Elsayed Radwan
Department of Computer Science,
Faculty of Computer and
Information Sciences,
Mansoura University, Egypt,
P.O.BOX:35516
elsfradwan@yahoo.com

***Abstract*— In this paper, we present a combinational approach for texture classification. The proposed method analyzes texture by 2D Discrete Wavelet Transforms (DWT); wavelet energy and some statistical features construct the features vector that characterizes texture. For improving accuracy the Probabilistic Neural Network (PNN), which is considered as a good estimator to the probability density function, is used as a classifier that maps input features vectors to the most appropriate texture classes. Two comparative evaluations have been done in order to ensure the effectiveness and efficiency of this model.**

Keywords- Texture classification, feature extraction, discrete wavelet transform, probabilistic neural network

I. INTRODUCTION

Texture is the variation of data at scales smaller than the scales of interest [7]. Techniques for the analysis of texture in digital images are essential to a range of applications in areas as diverse as robotics, medicine and the geo-sciences. In biological vision, texture is an important cue allowing humans to discriminate objects. This is because the brain is able to decipher important variations in data at scales smaller than those of the viewed objects. Texture may be important as well in object recognition as it tells us something about the material from which the object is made. In order to deal with texture in digital data, many techniques have been developed by image processing researchers [14] [7].

Texture classification aims to assign texture labels to unknown textures, according to training samples and classification rules by finding the best matched category for the given texture among existing textures. Two major issues are critical for texture classification: the texture feature extraction and texture classification algorithms [4]**.**

Texture feature extraction is considered as the main base of the efficiency of the texture classification algorithm. In order to design an effective algorithm for texture classification, it is essential to find a set of texture features with good discriminating power. Unfortunately, because of scale dependency of texture, its feature extraction has become a difficult problem. There have been many studies in solving texture classification problem based on various types of features and different methods of feature extraction. Most of the textural features are generally obtained from the application of a local operator, statistical analysis, or measurement in a transformed domain [14]. Generally, the features are estimated from Law's texture energy measures, Markov random field models, Gibbs distribution models and local linear transforms were found not to be robust enough to allow one-to-one mapping between patterns and parameter sets for many reasons: the parameters computed rely on the model assumed, the neighborhoods used must not be self-contradictory and they rely on the number of samples available for each combination of neighborhoods. In short, no model fits the observed textures perfectly, and so no model parameters are perfect in capturing all characteristics of a texture image [7][8]. Other studies used Fourier transform domain, fractals and co-occurrence matrices. The co-occurrence features such as contrast, homogeneity etc., were found to be the best of these features that they are popular due to the perceptual meaning they have. However, they are not adequate for texture and object discrimination as they throw away most of the information conveyed by the co-occurrence matrices [7].

In the recent years, wavelet analysis has become a powerful tool for multi-resolution analysis. Discrete Wavelet transform (DWT) and Gabor Transform are extensively used for texture analysis. While the DWT uses fixed filter parameters for image decomposition across scales, the Gabor Transform requires proper tuning of filter parameters for different scales of decomposition. Further, Wavelet based methods are shown to be efficient in detection, classification and segmentation for many reasons: the wavelet transform is able to de-correlate the data and achieve the same goal as the linear transformation, it provides orientation sensitive information which is essential in texture analysis and the computational complexity is significantly reduced by considering the wavelet decomposition [11][14].

As denoted before the efficiency of any classification system depends on effective characterization as well as

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

choosing the appropriate classifier. Some classifier algorithms such as support vector machines are used in some works which faced some problems because of high algorithmic complexity and extensive memory requirements [16]; the distance classifier is also used for measurement of similarity and consequent labeling but it suffered from some limitation in speed and adding parameters may cause the classifier to fail. [4].

In this study, a hybrid model based on the combinational approach is proposed, which combine 2D Discrete Wavelet Transform (DWT) and Probabilistic Neural Network (PNN) for solving texture classification problem. In the hybrid configuration, the 2D DWT is used for texture analysis and constructing features vector that characterizes the texture image by capturing all essential information. The obtained features vectors are then fed into the PNN which is used as a good estimator to probability density function that help in mapping each texture feature vector to the best appropriate class with fast and efficient performance. For illustrating the effectiveness of this model, two comparative evaluations have been done. The first one was among variety of wavelet filters for finding the best features extractor that provides the best characterization. The other was between the PNN and Backpropagation Neural Network (NN) as a classifier according to the mean success rates.

This paper is organized as follows; in section II, Discrete Wavelet transform (DWT), Probabilistic Neural Network (PNN) and Wavelet energy are mentioned. The hybrid model of 2D DWT and PNN is described in section III. The effectiveness of the proposed hybrid model for classification of texture images and comparative evaluations are demonstrated in section IV .Finally, section V presents discussion and conclusion.

II. PRELIMINARIES

A. Discrete wavelet transform

Wavelets are functions that satisfy certain mathematical requirements. They are used to cut up data into different frequency components and then study each component with a resolution matched to its scale. The basic idea of the wavelet transform is to represent any arbitrary function as a superposition of wavelets. Any such superposition decomposes the given function into different scale levels where each level is further decomposed with a resolution adapted to that level [12].

By applying DWT, the image is actually divided i.e., decomposed into four sub-bands and critically sub sampled as shown in Figure 1. (a). These four sub-bands arise from separable applications of vertical and horizontal filters. The sub-bands labeled LH1, HL1 and HH1 represent the finest scale wavelet coefficients, i.e., detail images while the sub band LL1 corresponds to coarse level coefficients, i.e., approximation image. To obtain the next coarse level of wavelet coefficients, the sub band LL1 alone is further decomposed and critically sampled. This result in two- level wavelet decomposition as shown in Figure 1. (b).This process continues until some final scale is reached. The values or transformed coefficients in approximation and detail images (sub-band images) are the essential features, which are shown here as useful for texture analysis and discrimination. As textures have non-uniform pixel value variations, they can be characterized by the values in the sub-band images or their combinations or derived features from these bands [11].

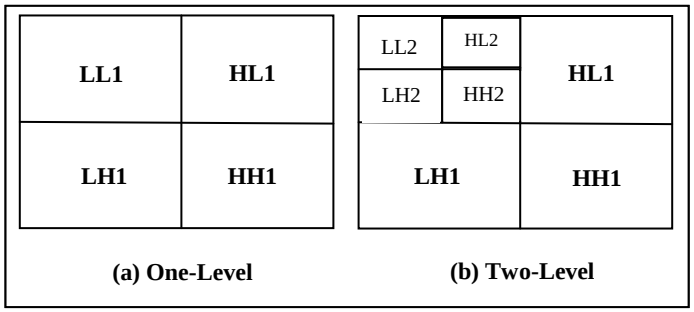


Figure 1. Image decomposition

B. Energy

Energy is one of the most commonly used features for texture analysis [4]. Wavelet energy reflects the distribution of energy along the frequency axis over scale and orientation and has proven to be very powerful for texture classification. The energy of sub-band C containing N coefficients is defined as in equation (1) [5],

$$Energy = \frac{1}{N} \cdot \sum_{i,j}^{N} \left(C(b_i, b_j) \right)^2 \quad (1)$$

C. Probabilistic neural network

It is shown that, by replacing the Sigmoid activation function often used in neural networks with an exponential function, a neural network can be formed which computes nonlinear decision boundaries. The resulted network is considered as an estimator to the probability density functions which can be used to map input patterns to output patterns and to classify patterns. This technique yields decision surfaces which approach the Bayes optimal under certain conditions [3].

PNN is a kind of these networks that called radial basis network. It is an artificial neural network with radial basis function (RBF) as a transfer function. RBF is a bell shape function that scales variable nonlinearly [15]. This network provides a general solution to pattern classification problems by following an approach developed in statistics, called Bayesian classifiers [6]. PNN is suitable for these kinds of classification problems for many advantages: Its training speed is many times faster than standard feed forward backprobagation network, it can approach a Bayes optimal result under certain easily met conditions and it is robust to noise examples.

The most important advantage of PNN is that training is easy and instantaneous that weights are not "trained" but assigned. Existing weights will never be alternated but only new vectors are inserted into weight matrices when training. So, it can be used in real-time. Since the training and running

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

procedure can be implemented by matrix manipulation, the speed of PNN is very fast [15].

The probabilistic neural network uses a supervised training set to develop distribution functions within a pattern layer. These functions, in the recall mode, are used to estimate the likelihood of an input feature vector being part of a learned category, or class. The learned patterns can also be combined, or weighted, with the a priori probability, also called the relative frequency, of each category to determine the most likely class for a given input vector. If the relative frequency of the categories is unknown, then all categories can be assumed to be equally likely and the determination of category is solely based on the closeness of the input feature vector to the distribution function of a class [6][1].

Probabilistic neural networks can be used for classification problems. When an input is presented, the first layer computes distances from the input vector to the training input vectors and produces a vector whose elements indicate how close the input is to a training input. The second layer sums these contributions for each class of inputs to produce as its net output a vector of probabilities. Finally, a competed transfer function on the output of the second layer picks the maximum of these probabilities, and produces a 1 for that class and a 0 for the other classes. The architecture for this system is shown below in Figure 2. [6].
It is assumed that there are Q input vector/target vector pairs. Each target vector has K elements. One of these elements is 1 and the rest are 0. Thus, each input vector is associated with one of K classes.
The first-layer input weights IW_1 are set to the transpose of the matrix formed from the Q training pairs, P'. When an input is presented, the $||\ dist\ ||$ box produces a vector whose elements indicate how close the input is to the vectors of the training set. These elements are multiplied, element by element, by the bias and sent to the radial basis transfer function. An input vector close to a training vector is represented by a number close to 1 in the output vector a_1. If an input is close to several training vectors of a single class, it is represented by several elements of a_1 that are close to 1.
The second-layer weights LW_2 are set to the matrix T of target vectors. Each vector has a 1 only in the row associated with that particular class of input, and 0's elsewhere. The multiplication Ta_1 sums the elements of a_1 due to each of the K input classes. Finally, the second-layer transfer function, compete, produces a 1 corresponding to the largest element of n_2, and 0's elsewhere. Thus, the network classifies the input vector into a specific K class because that class has the maximum probability of being correct [9].

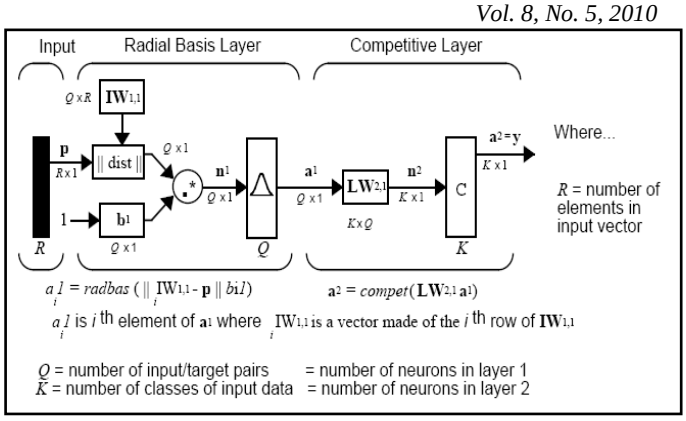


Figure 2. PNN Architecture

The bias b allows the sensitivity of the radial basis neuron to be adjusted. Each bias in the first layer is set to 0.8326/SPREAD. This determines the width of an area in the input space to which each neuron responds. SPREAD should be large enough that neurons respond strongly to overlapping regions of the input space [13].

The Probabilistic Neural Network is based on Bayesian classification and the estimation of probability density function that is necessary to classify the input vectors into one of the target classes approaching the Bayesian optimality [15].

III. HYBRID MODEL FOR TEXTURE CLASSIFICATION

The texture classification scheme is based on two principles, choosing features that provide the best characterization to the texture image and working with fast, easy and robust classifier in order to reach the best classification result.
In this study a viable algorithm with high precision and low calculating load is proposed to classify texture images using wavelet transform and its combination with probabilistic neural network. In the proposed combinatory configuration the DWT and PNN function as black boxes in a complementary manner. The functionality manner involved can be combined in two phases:

i. Texture characterization phase and
ii. PNN classification phase.

The texture classification phase starts with taking the texture images as an input and with the help of the DWT, the texture images are analyzed and features vectors are constructed.
The obtained features vectors are entered to the PNN for training which starting the PNN classification phase that continues with testing and ends with displaying the classification result as illustrated in Figure 3.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

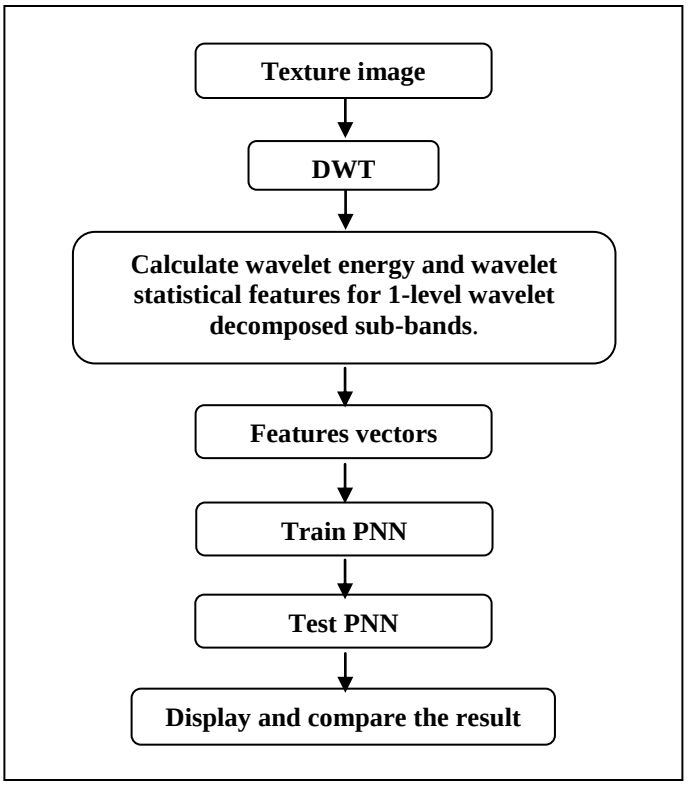


Figure 3. Structure of Hybrid Classification model

A. *Texture characterization phase*

In order to overcome the obstacle of texture characterization due to its scale dependent property, the discrete wavelet transform is used as a powerful tool for multi-resolution analysis.

For wavelet decomposition of various texture images, the decomposition at 1-level is performed using different wavelet transform filters. Thus, the image is decomposed into one approximate image with one approximation coefficients LL and three detail images with horizontal LH, vertical HL and diagonal HH detail coefficients.

Wavelet transform of an image measures light fluctuation in different scales. Therefore, the wavelet energy that reflects the distribution of energy along the frequency axis over scale and orientation is calculated for the approximation and the detail coefficients matrices. Also to increase sensitivity and precision, some wavelet statistical features are calculated such as the mean and the standard deviation of the approximation matrix as well as the mean and the standard deviation of the detail coefficients matrices; and then they are added to image features. The features obtained construct a feature vector with 12 elements organized as illustrated in Figure 4. The first four elements represent wavelet energy of the approximation and the detail coefficients matrices which is computed as in equation (1).

The second four elements represent the arithmetic mean of the approximation and the detail coefficients matrices which calculate the standard average of matrix elements that is computed as in equation (2),

$$\bar{x} = \frac{1}{N} \cdot \sum_{i,j}^{N} |C(b_i, b_j)| \quad (2)$$

And the last four elements represent the standard deviation of the approximation and the detail coefficients matrices which measure the variability in matrix elements and are computed as the square root of variance as in equation (3),

$$S = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \quad (3)$$

Wavelet Energy				Mean				Standard Deviation			
LL	*LH*	*HL*	*HH*	*LL*	*LH*	*HL*	*HH*	*LL*	*LH*	*HL*	*HH*

Figure 4. Features vector

Hence, at the end of this phase, each texture image has been well characterized by 12 elements features vector that captures all essential information needed for discrimination.

B. *PNN classification phase*

PNN is considered as a good estimator to probability density function that can be used to map input patterns to output patterns and to classify patterns efficiently with fast execution and ease implementation.

PNN classification phase has two parts, training and testing. After constructing the features vectors that represent the texture images in the texture characterization phase, the network is trained with the features vectors and the corresponding texture images as described in the following steps:

1) The input is of size $R \times Q$ with feature elements $R = 12$ and training samples$Q$.
2) Radial basis layer weight IW_1 is set to the transpose of the $R \times Q$ matrix of training samples that IW_1 is of size$Q \times 12$.
3) The dot product between the input vector p with size $12 \times 1$and the i^{th} row of IW_1 produces the i^{th} element of the distance vector $||IW_1 - p||$ whose size$Q \times 1$.
4) The radial basis layer biases b are all set to$0.8326 / SPREAD$, that $SPREAD$ is a constant chosen according to experiment.
5) The net input n_1 is obtained from element-by-element multiplication of the bias vector b with the distance vector $||IW_1 - p||$that denoted as

$$n_1 = ||IW_1 - p|| \cdot * b \quad (4)$$

6) The transfer function is the radial basis function that defined as in equation (5) and its shape is illustrated in Figure 5.

$$a = radbas(n) = e^{-n^2} \quad (5)$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

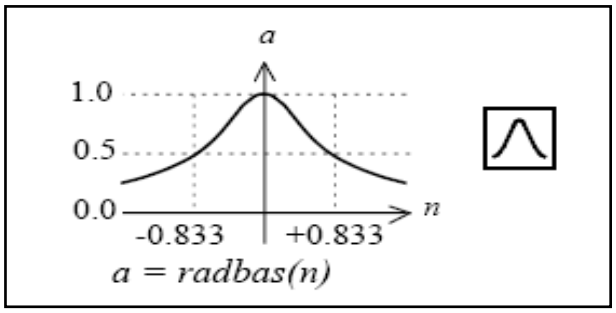


Figure 5. Radial basis function

If the input p is identical to the i^{th} row of IW_1, then the i^{th} element of a_1 is equal to 1
If the input p is close to the i^{th} row of IW_1, then the radial basis function produces a value near 1, else it produces a value far from 1.

7) Competitive layer weights matrix LW_2 is set to $K \times Q$ matrix with target classes$K$.
8) In competitive layer, the vector a_1 is multiplied with matrix LW_2 producing the output vector n_2 of size $K \times 1$.
9) The competitive function C produces 1 corresponding to the largest value of n_2 and 0 elsewhere.
10) For testing the network, an unknown features vector is entered as input and the network classifies it according to the class associated with the largest probability.

IV. EXPERIMENTAL RESULT

In order to assess the discrimination capability of the proposed hybrid model, experiments are conducted with 14 Vision Texture (Vistex) images database, each of size 512×512 which is divided into sub-images of size 64×64 for totally 896 texture images with 14 texture classes.
First, for applying texture characterization phase, one-level wavelet decomposition is applied to each texture image using *db2, sym6, db5, bior2.2, coif2, db3* wavelet filters separately.
After extracting the features, each texture image is characterized and represented with 12 elements features vector constructing totally 896×12 features vectors for all texture images used. Then, before entering the second phase, the PNN classification phase, the features vectors obtained at the end of the texture characterization phase are divided into 756 features vectors for training and 140 for testing.
Finally, the PNN is trained with 756×12 input features vectors and 14 target classes. After training, the PNN is tested with 140×12 features vectors.
For illustrating the effectiveness and efficiency of this model, two comparative evaluations have been done. The first one was between features extracted by the chosen wavelet filters according to the corresponding success rate in order to evaluate the efficiency of characterization for each filter. And the other was between the PNN and Backpropagation NN as a classifier according to the mean success rates for each classifier. The obtained results are discussed below.

The number of incorrect classifications of the PNN to the tested texture images using the wavelet filters mentioned before is tabulated in TABLE I. It is found that the wavelet filters *db2, bior2.2* and *db3* give the best characterization to texture images *Bark.0000* and *Paintings.11.0003*, wavelet filters *sym6, bior2.2* and *coif2* give the best characterization to *leaves.0008*, wavelet filters *db2, sym6* and *bior2.2* give the best characterization to *Metal.0004* and wavelet filters *sym6, db5, coif2* and *db3* give the best characterization to *Flowers.0006*.

TABLE I.THE NUMBER OF INCORRECT CLASSIFICATIONS WITH VARIOUS WAVELET FEATURES

No	Texture images	Number of incorrect classifications					
		(db2)	*(sym6)*	*(db5)*	*(bior2.2)*	*(coif2)*	*(db3)*
1	***Bark.0000.pgm***	**1**	**3**	**2**	**1**	**3**	**1**
2	***Bark.0006.pgm***	**1**	**1**	**1**	**1**	**1**	**1**
3	***Fabric.0015.pgm***	**1**	**1**	**1**	**1**	**1**	**1**
4	***Flowers.0006.pgm***	**2**	**1**	**1**	**2**	**1**	**1**
5	***Food.0000.pgm***	**0**	**0**	**0**	**0**	**0**	**0**
6	***Leaves.0008.pgm***	**1**	**0**	**1**	**0**	**0**	**2**
7	***Metal.0002.pgm***	**0**	**0**	**0**	**0**	**0**	**0**
8	***Metal.0004.pgm***	**0**	**0**	**1**	**0**	**1**	**2**
9	***Misc.0003.pgm***	**1**	**1**	**1**	**1**	**1**	**1**
10	***Paintings.11.0003.pgm***	**2**	**4**	**3**	**2**	**4**	**2**
11	***Water.0001.pgm***	**0**	**0**	**0**	**0**	**0**	**0**
12	***Stone.0004.pgm***	**1**	**1**	**1**	**1**	**1**	**1**
13	***Wood.0002.pgm***	**0**	**0**	**0**	**0**	**0**	**0**
14	***Water.0005.pgm***	**0**	**0**	**0**	**0**	**0**	**0**

The mean success rates of the chosen wavelet filters are shown in Figure 6. It is found that the characterization of texture images using *bior2.2* wavelet filter has the highest mean success rate (93.57%) and the lowest mean success rate (90.72%) was to the characterization using *coif2* wavelet filter which is observed in Figure 7.

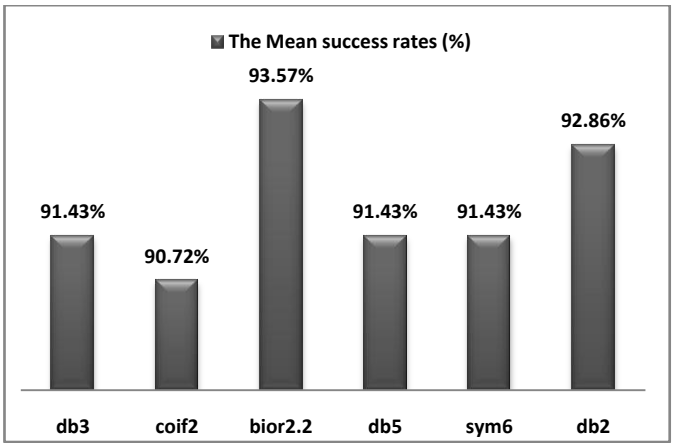


Figure 6. The mean success rates of various wavelet filters

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

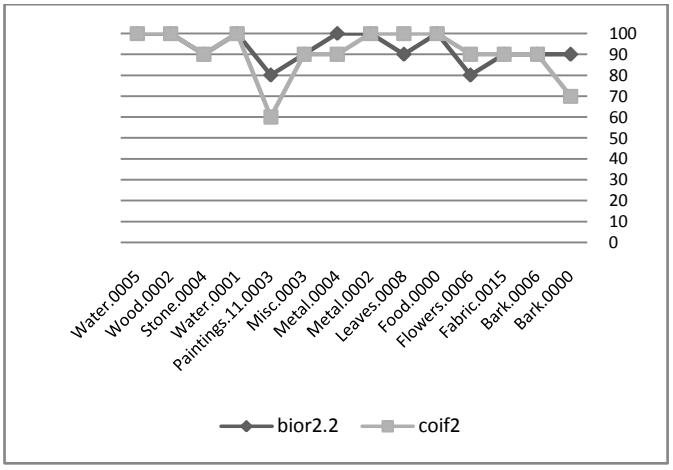


Figure 7. Comparison between the best and the worst characterization according to the corresponding correct classification rates of 14 texture classes.

The features vectors that have been obtained from *bior2.2* wavelet filter which represent the best characterization of the texture images have been used to compare the performance of the PNN used in this model and the Backpropagation NN in classification.

As illustrated in Figure 8, the mean success rate that has been obtained over the 14 texture classes using the PNN is far more than the other obtained using the Backpropagation NN. In addition to the great differences in speed and simplicity of the performance between the two classifiers prove the effectiveness and efficiency of the PNN that has been used in this model with respect to Backprobagation NN.

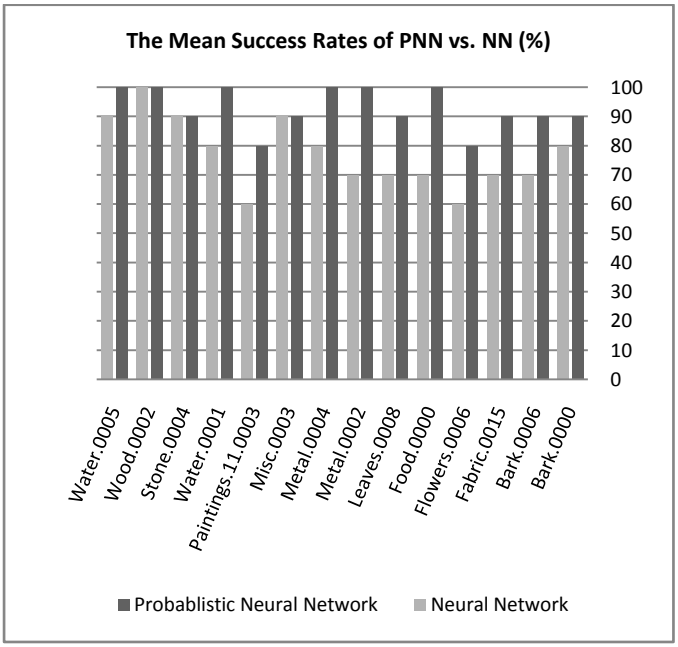


Figure 8. Comparison between the mean success rates that have been obtained using the PNN verses the mean success rates that have been obtained using Backpropagation NN.

V. CONCLUSION

Texture classification problem has become the business of a lot due to its great effect on many fields. As an attempt to solve this problem many studies have been done in spite of hindrances they suffered in the two sides, characterization and classification. As for characterization, researchers tried to choose features that provide the best representation of the texture images in the presence of texture scale dependent property. While they work on the use of the appropriate classifier algorithm to provide the best discrimination capability.

In this work a hybrid model is presented to classify texture images. The 2D DWT is combined with the PNN constructing this hybrid model. The discrete wavelet transform is used as a powerful tool for multi-resolution analysis so it is used for texture analysis as an attempt to overcome the obstacle of texture scale dependent property. The PNN is a radial basis network that is considered as an estimator to the probability density functions which can be used to map input patterns to output patterns and to classify patterns. The PNN is suitable for these kinds of classification problems that it can approach a Bayes optimal result under certain easily met conditions; as well as the training is easy, fast and robust.

The structure of the proposed hybrid model is divided into two phases. The first phase is the texture classification phase in which 1-level wavelet decomposition has been performed and 4 sub-bands have been obtained representing the approximation, vertical, horizontal and diagonal detail. As a way to select features that capture all the essential information needed to uniquely characterize the texture, wavelet energy that reflects the distribution of energy along the frequency axis over scale and orientation has been calculated for the approximation and the detail coefficients matrices. Also to increase sensitivity and precision, some wavelet statistical features are calculated such as the mean and the standard deviation of the approximation matrix as well as the mean and the standard deviation of the detail coefficients matrices; and then they are added to image features. The features obtained construct a feature vector with 12 elements which is fed with the corresponding target class as input to the PNN starting the second phase, the PNN classification phase. In the PNN classification phase, the PNN is trained with the input features vectors then it is tested with other features vectors for evaluating its discrimination capability.

Experiments have been conducted for evaluating the performance of the proposed hybrid model. The model proved that the features derived from the approximation and detail coefficients, the wavelet energy and the statistical features, uniquely characterize a texture. In order to find the best features extractor a comparative evaluation has been done with features extracted by different wavelet filters and the corresponding correct classification rates. Another comparative evaluation with respect to classifiers has been performed between the PNN and Backpropagation NN which provided great evidence about the effectiveness of the PNN as

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

a classifier for texture images with simple structure, fast execution and high performance.

This hybrid classification model achieved a good progress in solving the texture classification problem by the proposed complementary manner of the 2D DWT and PPN. Compared with other methods, the system functionality is fast in execution, efficient in recognition and easy in implementation.

We are still in a need of finding more features that capture more essential information of texture in order to provide the best characterization and achieve the optimal classification results. More extended efforts are under development in order to improve the efficiency of the system.

ACKNOWLEDGMENT

First of all, I thank my father, my mother, my sisters and my brother for their moral support I required in my life at all.

I deeply thank my advisor, Dr. ElSayed Radwan, whose help, advice and supervision was invaluable.

Lastly, I offer my regards and blessings to my friends and all of those who supported me in any respect during the completion of the paper.

REFERENCES

[1] Ali H. Al-Timemy, Fawzi M. Al-Naima and Nebras H. Qaeeb, "Probabilistic Neural Network for Breast Biopsy Classification," MASAUM Journal of Computing, Iraq, September 2009.

[2] Anagnostopoulos I., Anagnostopoulos C., Vergados D., Loumos V., and Kayafas E., "A Probabilistic Neural Network for Human Face Identification based on Fuzzy Logic Chromatic Rules," IEEE MED03, Greece, 2007.

[3] Donald F. Specht, "Probabilistic neural networks for classification, mapping, or associative memory," IEEE International Conference on Neural Networks, California , 2001.

[4] Engin Avci, Abdulkadir Sengur and Davut Hanbay, "An optimum feature extraction method for texture classification," Pergamon Press, Inc. Tarrytown, NY, USA, 2009.

[5] G.Van de Wouwer, P.Scheunders and D.Van Dyck, "Statistical texture characterization from discrete wavelet representations," IEEE Transactions on Image Processing, Belgium,1999

[6] Leila Fallah Araghi, Hamid Khaloozade and Mohammad Reza Arvan, "Ship Identification Using Probabilistic Neural Networks (PNN)," The International Multi-Conference of Engineers and Computer Scientists, Hong Kong, March 2009

[7] Maria Petrou and Pedro Garcia Sevilla, "Image processing-Dealing with texture," John Wiley&Sons Ltd, England, 2006

[8] M.Ghazvini, S.A.Monadjemi, N.Movahhedinia and K.Jamshidi, "Defect Detection of Tiles Using 2D-Wavelet Transform and Statistical Features, " World Academy of Science,Engineering and Technology, Iran, 2009

[9] Neural Network Matlab Toolbox 2008

[10] Sami Gazzah and Najoua Essoukri Ben Amara, "Arabic Handwriting Texture Analysis for Writer Identification Using the DWT-Lifting Scheme, " Ninth International Conference on Document Analysis and Recognition, Parana, 2007

[11] S.Arivazhagan and L. Ganesan "Texture Classification using Wavelet Transform ," The Sixth International Conference on Computational Intelligence and Multimedia Applications (ICCIMA'05), IEEE, India, 2005

[12] S.Arivazhagan, S.Deivalakshmi and K.Kannan, " Performance Analysis of Image Denoising System for different levels of Wavelet decomposition," International Journal of Imaging Science and Engineering (IJISE), USA, JULY 2007

[13] S.N. Sivandm, S. Sumathi and S.N.Deepa, "Introduction to Neural Netwrks Using Matlab7.0," Tata mcgraw-Hill Publishing Company Limited, Copyright 2006.

[14] S.Shivashankar and P.S.Hiremath, "Wavelet Based Features For Texture Classification," ICGST International Journal on Graphics, Vision and Image Processing, India, December 2006

[15] Stephen Gang Wu, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chang and Qiao-Liang Xiang, "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network," arXiv, USA, July 2007.

[16] http://www.svms.org/disadvantages.html, 2010-07-07

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A New Learning Method for Cellular Neural Networks Templates based on Hybrid of Rough Sets and Genetic Algorithms

Elsayed Radwan *, Omaima Nomir
Department of Computer Science, Faculty of CIS,
Mansoura University, Egypt
E-mails: *elsfadwan@yahoo.com, o.nomir@umiami.edu*

Eiichiro Tazaki
Department of Control and Systems Engineering, Toin
University of Yokohama, Japan
E-mail: *tazaki@intlab.toin.ac.jp*

Abstract— **A simple method for synthesizing and optimizing Cellular Neural Networks is proposed. Based on the Rough Sets concept and the comparison principles for ordinary differential equations, a mathematical system of inequalities and the optimal cloning template structure are discovered. By solving this system of inequalities, the derived parameters are represented to be the Cellular Neural Networks templates. These parameters guarantee correct operations of the network. To represent a more robust template, a randomized search and an optimization technique guided by the principles of evolution and nature genetics with constrained fitness and, penalty functions, has been introduced. Applying our introduced method to different applications shows that our new method is robust.**

Keywords-component; Rough Sets; Cellular Neural Networks, Comparison principles; Template Robustness; Genetic Algorithms

I. INTRODUCTION

Cellular Neural Networks [2], CNN were invented to circumvent this curse of interconnecting wires. The problem gained by the fully connected Hopfield Network [10], is by decreasing that and there should be no electrical interconnections beyond a prescribed sphere of influence. This makes it easy to be implemented via physics device as VLSI (Very Large Scale Integrated) Circuit. During the CNN invention period, due to the lack of any programmable analogic CNN chips, the templates were designed to be operational on ideal CNN structures. These structures were simulated on digital computers. Later, several templates learning and optimization methods were developed. The goal of these methods was template generation, dealing with ideal CNN behavior but without much regard to robustness issues. As a result, a large number of templates were introduced. Some of these templates were designed by using template learning algorithms, but most of them were created relaying on ad hoc methods and intuition. Since the programmable CNN chips were fabricated, many of these templates were found to work incorrectly in their original form (i.e. as used in software simulators). Consequently, new chip- independent robust template's design methods were introduced. According to previous studies [8], the actual template values at each cell will be different from the ideal ones. This is mainly due to the noise in the electrical components, superfluous cells as well as the template parameters. This results in some cells responding erroneously to some inputs. An improvement

should be achieved by designing robust templates for a given CNN's operations so that they are most tolerant against parameter deviations. This can be achieved by removing the cells that have no effect on classifying the output, and superfluous cells, removing noise in the training data, and discovering the optimal template parameters.

In this paper, we introduce an analytical method to synthesize a CNN for solving a given problem. Our introduced method relies on Rough sets concepts [15] in discovering the optimal template structure by removing the superfluous neighboring cells which have no effect on classifying the cell's output. Another important concept of rough sets is its ability to determine the significance of each neighbor cell. This rough sets' feature gives us the idea to define a new measure called the sign measure. This measure is used in deducing the relation among the template parameters. Also, by rough set concepts the similarities in the input data are discovered and excluded, which will result in reducing the learning time. Moreover, it is able to discover the optimal local rules of the most simplified construction, which (almost) preserve consistency with data and classify so far unseen objects with the lowest risk of error. Therefore the capability of classifying more objects with high accuracy, increase the CNN template robustness, and that needs neglecting cells being the source of redundant information. Depending on the local rules, our method uses a simple procedure of the so-called comparison principle [3], which provides bounds on the state and output waveforms of an analog processing cell circuit. We will be able to find conditions on the elements of the CNN, ensuring a correct functioning of the CNN for a particular application. To find the global minima, even in a noisy and discontinuous search space and without using differentiable information about the cost function, Genetic Algorithms with constrained fitness function [17] that takes into account the hardware implementation is used. This research work is an extension of the previous work [6], where a special case of CNN is handled. Rough sets are used in discovering the optimal CNN template structure. Also, the comparison principle technique is used to treat the regular discovered rough sets' rules to be a set of inequalities that constraints the CNN structure. The problem of uncoupled CNN in designing a simple application of edge-detection CNN. is solved.

The rest of this paper is organized as follows: Section 2 explains the role of rough set concepts in reasoning about cells and concludes the optimal local rules that describe the CNN dynamic. Section 3, describes the Genetic algorithm in learning the cloning templates. Sections 4 presents the experimental results on some simple applications and then section 5 concludes the paper.

II. ROUGH SETS IN REASONING ABOUT CELLS

Cellular Neural Networks [4] is any spatial arrangement of nonlinear analogue dynamic processors called cells. Each cell interacts directly within finite local neighbors that belong to the sphere of influence $N_r(ij) = \{c_{kl} \mid \max(|i-k|,|j-l|) \le r\}$ and characterized by a nonlinear dynamical system. This dynamical system has an input u, a state x evolved by time according to some prescribed dynamical laws, and an output y, which is a function of the state. The cell dynamics are functionally determined by a small set of parameters which control the cell interconnection strength called templates. It is characterized by the following equations [1].

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$C\frac{d}{dt}x_{ij}(t) = -R^{-1}x_{ij}(t) + \sum_{kl\in N_{r(i,j)}} A_{ij,kl}y_{kl}(t) + \sum_{kl\in N_r(ij)} B_{ij,kl}u_{kl}(t) + z \quad (1)$$

$$y_{ij}(t) = f(x_{ij}(t)) = 0.5(|x_{ij}(t)+1| - |x_{ij}(t)-1|) \quad (2)$$

$$-1 \le x_{ij}(0) \le 1, \quad -1 \le u_{ij}(t) \le 1, \quad |z| \le z_{\max}, \quad 1 \le i \le M, \quad 1 \le j \le N \quad (3)$$

A and B are the feedback and the feed-forward templates respectively, and z is the search bias threshold. The machine uses the simple CNN in a time-multiplexed fashion, analogous to the ALU of the microprocessor, by controlling the template weights and the source of the data inputs for each operation. The machine supplies memory and register transfers at each cell that allow the outputs of the CNN operations to be combined and/or supplied to the inputs of the next operations, thereby allowing more complex algorithms to be implemented. Then, for any input pattern U, the output for each cell $y_{ij}(\infty)$ is uniquely determined by only a small part of U, depicted in Figure 1 where the radius of the sphere of influence $r = 1$, exposed to $(2r+1)\times(2r+1)$ transparent window centered at cell C_{ij}. According to the complete stability theorem of the uncoupled CNN [1] [5], the output $y_{ij}(\infty)$ is considered as a function in $(2r+1)(2r+1)$ of input variables in addition to a predefined initial state x_0, $y_{ij} = f(x_0, u_1, ..., u_{(2r+1)(2r+1)})$. The functionality of the uncoupled CNN is a one-one mapping from U to Y for a predefined initial state x_0 that describe the dynamic at $t = 0$.

Hence, the dynamic for space invariant uncoupled CNN can be completely described by a Knowledge Representation System, *KRS*, $S = (U, X_0 \cup C \cup Y)$ where U is the whole universe of input pattern and C is the neighbor cells, Y is the output from a predefined initial state X_0, $Y \notin C$ [4]. Then, every row h in S is considered as an if-then rule by the form;

$$if\ ((c_0 = x^h(t_0)) \,\&\, (c_1 = u_1^h) \,\&\, ... \,\&\, (c_5 = u_5^h) \,\&\, ... \,\&\, (c_9 = u_9^h))\ \ then\ y = y^h \quad (4)$$

To summarize, it can be described as a CY decision rule $\phi \to \psi$ where the predecessor ϕ is a conjunction of $(2r+1)(2r+1)+1$ of input cells, (c_i, u_i^h), and the successor ψ is the classified output. This means, the whole *KRS* looks like a collection of CY decision rules or in short CY decision algorithm, $Dec(C,Y) = \{\phi_k \to \psi_k\}_{k=1}^m, 2 \le m \le |U|$. This decision algorithm can be treated by an algorithm for synthesis of decision rule from decision table. Rough Sets [12][15] provides a mathematical technique for discovering the regularities in data, which aims to obtain the most essential part of the knowledge that constitutes the reduced set of the system, called the reduct. It depends on the analysis of limits of discernibility of subsets of objects from the universe of discourse *U*. For this reason, it introduces two subsets, the lower and upper approximation sets. With every subset of attribute $\overline{C} \subseteq C$, any equivalence relation $I_{\overline{C}}$ on *U* can easily be associated to;

$$I_{\overline{C}} = \{(x,y) \in U : \text{for every } c_i \in \overline{C}, c_i(x) = c_i(y)\} \quad (5)$$

Then $I_{\overline{C}} = \cap_{c_i \in \overline{C}} I_{c_i}$. If $X \subseteq U$ (2), the sets $\{x \in U : [x]_{\overline{C}} \subseteq X\}$ and $\{x \in U : [x]_{\overline{C}} \cap X \neq \phi\}$ (3) where $[x]_{\overline{C}}$ denotes the equivalence class of the object $x \in U$ relative to $I_{\overline{C}}$, which are called the $\overline{C}$ *-lower* and $\overline{C}$ *-upper* approximation of X in *S*. Through this paper, rough set relies on discovering the consistency relation among the rules, by means of decision language, and determining the dependencies among data. The rules of the most simplified construction, (almost) preserve consistency with data, are likely to classify so far unseen object with the lowest risk of error. Therefore, to be capable of classifying more objects with high accuracy, we need to neglect cells being the source of redundant information, i.e. use the reduct of attributes.

Definition 1: if for every $\phi \to \psi$, $\phi' \to \psi' \in Dec(C,Y)$, we have $\phi = \phi'$ implies $\psi = \psi'$, then $Dec(C,Y)$ is called consistent algorithm, otherwise it's called inconsistent algorithm. Also we defined the positive region of $Dec(C,Y)$ denoted $POS(C,Y)$ to be the set of all consistent rules in the algorithm.

A cell attribute $c_i \in C$ is dispensable (superfluous) in $Dec(C,Y)$ if $POS(C,Y) = POS(C - \{c_i\}, Y)$; otherwise the cell attribute $c_i \in C$ is indispensable in $Dec(C,Y)$. The algorithm $Dec(C,Y)$ is said to be independent if all $c_i \in C$ are indispensable in $Dec(C,Y)$.

The set of cell attributes $\overline{C} \subseteq C$ will be called a reduct of $Dec(C,Y)$, if $Dec(\overline{C},Y)$ is independent and $POS(C,Y) = POS(\overline{C},Y)$. Based on the significance of each cell, the algorithm to compute the reduct is as follows;

1- Let $R = \varphi$, $C = \{c_0, c_1, c_2, \ldots, c_{(2r+1)(2r+1)}\}$ and $i = 0$

2- Compute the accuracy measure of the original table $k = \dfrac{|POS(C,Y)|}{|Dec(C,Y)|}$

3- While $(i <= (2r+1)(2r+1))$ do

a- Compute the accuracy measure by dropping the cell C_i, $k_i = \dfrac{|POS\ (C - \{c_i\}, Y)|}{|Dec\ (C,Y)|}$

b- If ($\gamma_{c_i} = k - k_i = 0$)

- Let $R = R \cup \{c_i\}$ and

- $C = C - \{c_i\}$

c- $i = i + 1$

γ_{c_i} is called the cell significance which represents the bifurcations in the CNN dynamical system caused by removing the cell c_i. If *k*

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

equals one, i.e. consistent algorithm, then the algorithm describes a complete stable dynamic. Suppose that the CNN dynamic could be represented by a decision algorithm and then study the affection of consistency relation in realizing the optimal template structure of a single layer CNN through the following theorems.

Theorem 1: Any consistent algorithm can be recognized by a CNN template for which, for all linear cells, there is no other direct connected linear cells, where a cell C_{ij} is directly connected to a cell C_{mn} if $|i-m| \le r$ and $|j-n| \le r$, i.e. $C_{mn} \in N_r(ij)$ and the feed-back synapse $A_{i-m,j-n} \neq 0$.

Proof: Let C_L be a linear cell, i.e. $x(t) = y(t)$. For any consistent algorithm, all cells that are directly connected to C_L must have constant output. Then the dynamics of $x_L(t)$ in the linear region governed by $\frac{dx_L}{dt} = x_L(t)(a_c - 1) + q$ where q comprises the contribution of the neighbor output values from the input, bias, and boundary which is constant by assumption as long as C_L is linear. $a_c = A_{00}$, where A_{00} is the center element of the A-template. The solution is a single exponential function with a positive argument, which guarantees that the equilibrium lies in the saturation region. Hence the sign of $\frac{dx_L(t)}{dt}$ determines the output values of the neighboring cells and can not change while the linear region. Therefore, the template is uncoupled CNN or there is no direct connected linear cells.

Corollary 1: Any inconsistent algorithm can not be realized by a single layer space invariant CNN without directly connected cells.

Proof: we prove that using contradiction by considering the opposite i.e. consider its inconsistent algorithm and can be represented by CNN with no directly connected cells. Then, the CNN dynamic can be represented by

$\frac{dX}{dt} = AX + W, A = (A_{00} - 1)I$ and $W = BU + z$, I is the identity matrix (6)

Then, $X = C_0 e^{(A_{00}-1)t}$ and $C_0 = C_0(A,W)$ is a linear function depends on a self-feedback constant value and the offset level which is a function of the input pattern. Hence the trajectory depends on anential function on time, i.e. it's a continuous monotonic function converges to a single equilibrium point. Thus, consider that $\phi \to \psi \in Dec(C,Y)$ and since ϕ based on the input pattern, C_0 is a constant value, and ψ is determined by a linear piecewise function in Equation (2) as function in the trajectory which converges to a single equilibrium point. Hence ψ is a one-one function which contradicts the definition of inconsistence. Therefore, we reject our assumption which completes the proof.

Since consistency of the algorithm gives no promise for the linearly separable, such as *XOR* logic function which is consistent but non-linearly separable, then the algorithm should be checked for linear separability. If it is a consistent algorithm and linearly separable, it can be recognized by a stable dynamic with memoryless parameters, and synaptic weights, according to the following theorem;

Theorem 2: Any consistent algorithm which is linearly separable and has c_0 as a superfluous attribute, it can be recognized by a single layer CNN with memoryless synaptic weight.

Proof:

Since CNN is a massively parallel architecture working with analog signals, and as the path of information is an analog to digital converter, then our proof will be concentrated on the binary output only.

Case 1: (binary input signals)

Since any consistent algorithm with binary signals can be seen as a truth table, then it can be determined by a statement form in which the only connectives occurring are from amongst, ($\sim, \wedge, \vee$, negation, conjunction and disjunction functions). Since for any local linearly separable Boolean function, there exists a barrier (plane) satisfies that the output at each cell $y = \text{sgn}[< a, x > -b]$. According to [1], they proved that any local Boolean function $\beta(x_1, x_2, ..., x_9)$ of nine variable is realized by every cell of an uncoupled CNN. This happens if and only if $\beta(.)$ can be expressed explicitly by the formula $\beta = \text{sgn}[< a, x > -b]$ where $< a, x >$ denoted the product between the vectors $a = [a_1, a_2, ..., a_9]$ and $x = [x_1, x_2, ..., x_9]$, where $a_i \in R, b \in R$ and $x_i \in \{-1, 1\}$ is the i[th] Boolean variable, i =1,…,9. Hence, there exists a single layer CNN with memoryless synaptic weights that realize the output, which satisfy the proof.

Case 2: (analog input signals)

We prove by considering the opposite, i.e. the output can not be recognized by a single layer CNN. Thus, by a single layer there exists an error corresponding to some cells C_E. This means that some cells remains in the linear saturation region or in the opposite saturation region. From theorem 1, any cell in the CNN including C_E should be realized a template for which, for all linear cells, there is no other directly connected linear cells, which is completely stable dynamic, i.e. all cells should belong to only one of the positive or negative saturation region. Hence, C_E should be in opposite saturation region, this case should be happened when C_E located in one of the degenerate cases. From the assumption about the binary output, there is only one degenerate case when the self-feed back A_{00} is greatest than one (i.e. CNN dynamic depends on the initial state), which contradicted with theorem 1 and c_0 is the superfluous attribute. This leads us to reject our assumption and the consistent algorithm is recognized by a single layer. Since the algorithm is linearly separable, the output can be recognized by a single layer CNN with memoryless synaptic weights.

After determining the set of reduct $\overline{C}$ and cells significance, we construct the CNN structure by removing the cells that corresponding to the attributes in the set R, the set of superfluous cells. Also, the decision table should be modified by removing the columns corresponding to the superfluous cells. Coupled to this, if c_0 belongs

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

to the set of reduct $\overline{C}$, i.e. the cell significance $\gamma_{c_0} \neq 0$, we can say that the output depends on the initial state, i.e. we should choose a strong positive self feedback weight $A_{00} > 1$ [4].

Corollary 2: Any consistent algorithm which is linearly separable can be recognized by a single layer Uncoupled CNN.

Proof: the proof comes as a direct result from theorem 1 and theorem 2.

Corollary 3: Every consistent local function of nine variables can be realized by ORing Uncoupled CNN.

Proof: This corollary is a direct result for the Min-term theorem [4] and theorem 1.

Since inconsistence of the algorithm that describes a dynamical system comes from noise in the handled data, this case is out of this paper scope, or from some activate cells that evolved by time, then there exists at least a direct connected linear cell that has its own effect on the center cell. This gives us the direction to expand our problem to handle the general case of the coupled CNN. To discover the optimal template structure of the coupled CNN, we will consider more constraints on the stability of the network. However, the stability of the CNN as a dynamical system gives a promise for a locally regular dynamic system. In regular dynamic system, A phase diagram for a given system may depend on the initial state of the system (as well as on a set of parameters), but often phase diagrams reveal that the system ends up doing the same motion for all initial states in a region around the motion, almost as though the system is attracted to that motion. Such attractive motion is fittingly called an attractor, a trajectory, for the system and is very common for forced dissipative systems.

Our model depends on considering more constraints for the stability so that the output of the neighboring cells around the attractors should have their effect on classifying the center cell's output. In inconsistence criteria, our model includes the output of some neighbor cells as additional attributes that able to classify the center cell output to deduce a modified decision table. This can be done by adding the output of the neighbors' active cells except the cell itself, wherever the cell output classify itself, that belong to the sphere of influence $\overline{N}_r(i,j)$, $(2r+1)(2r+1)$ of the cells that represent the desired output pattern, to the reduced cells $\overline{C}$. That is because of discovering the active output cells from the modified table. In the modified table, the set of attributes C will be expanded to be;

$$C = \overline{C} \cup \left\{ y_k \mid y_k \in \overline{N}_r(ij) / y_{ij}, k = N(i-1) + j \right\} \quad (7)$$

where its size $|C| = (2r+1)(2r+1) + |\overline{C}| - 1$. Based on the modified table, rough set concept will check the consistent rules. According to the consistency of the modified algorithm, we deduce that the number of layers and the optimal coupled CNN structure for increasing the radius to the sphere of influence in purpose of getting more attributes to classify the cell output. If the modified algorithm were still inconsistent, then we should add additional layers to represent the algorithm according to the later corollary;

Corollary 4: The modified algorithm that is inconsistent can not be recognized by a single layer CNN.

Proof: If we define

$$g_{ij}(t) = \sum_{C_{kl} \in N_{ij} \setminus \{C_{ij}\}} A_{k-i,l-j} y_{kl}(t) + \sum_{C_{kl} \in N_{ij}} B_{k-i,l-j} u_{kl} + z$$ We can restate the state equation of the coupled CNN as follows

$$\dot{x}_{ij} = -x_{ij}(t) + A_{00} y_{ij}(t) + g_{ij}(t)$$

By corollary 1, we prove that this can not be realized by single layer CNN.

Definition 2: The robustness of a template T denoted by $\rho(T)$ is defined as the minimal distance of the hyper-plane from the vertices of the hyper-cube.

Theorem 3: Let $F(u_1, u_2, ..., u_n)$ be an arbitrary n dimensional linearly separable function and π is the hyper plane separating the vertices. With decreasing the dimensionality from (n) to (n-1), the distance of vertices from π in (n-1) dimensions cannot be decreased.

Proof:

Let $V(v_1, v_2, ..., v_n)$ be an arbitrary vertex of the hypercube corresponding to the Function F, $\overline{w} = (w_1, w_2, ..., w_n)$ is the normal vector of π, $O = (o_1, o_2, ..., o_n) \in \pi$ such that $\overline{VO} \parallel \overline{w}$ ($\left|\overline{VO}\right|$ is the distance from V to π). If i is the dimension to be eliminated, for simplicity, we assume that $v_i = 0$. Let L be the projection of π onto (n-1)-dimensional hyper-cube corresponding to $F(u_1, u_2, ..., u_{i-1}, 0, u_{i+1}, ..., u_n)$, furthermore, $K(k_1, k_2, ..., k_{i-1}, k_{i+1}, ..., k_n) \in L$ such that

$\overline{VK} \parallel \left(\overline{w_1, w_2, ..., w_{i-1}, 0, w_{i+1}, ..., w_n}\right)$ ($\left|\overline{VK}\right|$ is the distance from V to L). The equations of L and π are as follows:

$\pi : w_1 u_1 + ... + w_i u_i + ... + w_n u_n + w_0 = 0$,

L :

$w_1 u_1 + ... + w_{i-1} u_{i-1} + w_{i+1} u_{i+1} + ... + w_n u_n + w_0 = 0$

Since $O = (o_1, o_2, ..., o_n) \in \pi$ and $K(k_1, k_2, ..., k_{i-1}, k_{i+1}, ..., k_n) \in L$

$$w_1 o_1 + ... + w_i o_i + ... + w_n o_n + w_0 = 0 \quad (8)$$

$$w_1 k_1 + ... + w_{i-1} k_{i-1} + w_{i+1} k_{i+1} + ... + w_n k_n + w_0 = 0 \quad (9)$$

Then, from (7) and (8) we have,

$$w_1 o_1 + ... + w_i o_i + ... + w_n o_n + w_0 = w_1 k_1 + ... + w_{i-1} k_{i-1} + w_{i+1} k_{i+1} + ... + w_n k_n + w_0 \quad (10)$$

This implies that,

$$w_1 o_1 + ... + w_i o_i + ... + w_n o_n = w_1 k_1 + ... + w_{i-1} k_{i-1} + w_{i+1} k_{i+1} + ... + w_n k_n \quad (11)$$

then

$$w_1(o_1 - k_1) + ... + w_{i-1}(o_{i-1} - k_{i-1}) + w_i o_i + w_{i+1}(o_{i+1} - k_{i+1}) + ... + w_n(o_n - k_n) = 0 \quad (12)$$

Hence, $\overline{KO} \perp \overline{w}$, since $\overline{VO} \parallel \overline{w}$ therefore $\overline{KO} \perp \overline{VO}$ using Pythagoras theorem,

$\left|\overline{VO} + \overline{OK}\right|^2 = \left|\overline{VO}\right|^2 + \left|\overline{OK}\right|^2$ this implies that

$\left|\overline{VK}\right|^2 = \left|\overline{VO}\right|^2 + \left|\overline{OK}\right|^2$, Since $\left|\overline{OK}\right|^2 \geq 0$, then,

$\left|\overline{VK}\right|^2 \geq \left|\overline{VO}\right|^2$, hence $\left|\overline{VK}\right| \geq \left|\overline{VO}\right|$. This completes the proof.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Corollary 5: The template's robustness caused by removing the superfluous cells is better than the robustness of the original template.
Proof: the proof comes as a direct result from theorem 3.
According to corollary 5, we can prove that by decreasing the number of effective cells by means of rough sets concepts, the template robustness should be improved.

- Speaking about cell attributes, it is obvious that they may have varying importance in the analysis of the issues being considered. This importance can be pre-assumed on the basis of auxiliary knowledge and expressed by property chosen's weights. Even though, our method relies on deducing the optimal template structure by discovering the optimal local rules, this method is not totally expressible for CNN with propagating type associated with gray inputs. This is because we reconstruct the modified table by taking the cells' output around the equilibrium points, i.e. expresses the output in the saturation region and away from the linear region. Therefore, a new measure should be discovered. We study the affection of cell attribute significance on determining the relation among the CNN template parameters. The cell significance $\gamma_{c_i} = k - k_i$ expresses how the positive region of the classification $U / IND(C)$ when classifying the object by means of cell attributes C will be affected when dropping the cell attribute c_i from the set C. In other words, $\gamma_{c_i} = k - k_i$ expresses the percent of local rules that are lost by dropping the cell attribute c_i. Also, it describes the relation between the input and the output when dropping the cell c_i by excluding the template i[th] cell attribute's parameter. Since the output is considered as a function of the cells input by defining a template $\Im$, $y = M(u, x_0 / \Im)$, then the cell significance should have its affect on describing the relation among cells strength, or CNN's template parameters.
- From our definition that CNN is an analog to digital converter, each local rule that describe the CNN dynamic should belong to only one positive rule set when the output is black or it should belong to negative rule set when the output is white. By considering the number of positive and negative rules, N^+ and N^- respectively, the probability of positive (negative) output is $P(y=1) = \frac{N^+}{N}$ $(P(y=-1) = \frac{N^-}{N}), N = N^+ + N^-$.
- Then the expected value to get positive output is $E(X^+) = N^+P(+) - N^-P(-) = N^+ - N^-$. By dropping a cell C_i from the reduced table, the positive and negative outputs should be disturbed as $N_{c_i}^+$, $N_{c_i}^+ \le N^+$, and $N_{c_i}^-$, $N_{c_i}^- \le N^-$, respectively, $N_{c_i} = N_{c_i}^+ + N_{c_i}^- < N$. The conditional probability of positive (negative) output by dropping c_i is $P(y = 1 / dropping \quad c_i) = \frac{N_{c_i}^+}{N_{c_i}}$ ($P(y = -1 / dropping \quad c_i) = \frac{N_{c_i}^-}{N_{c_i}}$). Then, the conditional expectation of getting positive output by dropping the cell c_i is $E(X_{c_i}^+ / dropping\ c_i) = N_{c_i}^+ - N_{c_i}^-$. Since the output for uncoupled CNN is completely determined by the following relation $y_\mu = \mathrm{sgn}[(A_{00} - 1)x_\mu(0) + w_\mu]$ therefore. the probability can be expressed as follows,
- $P((A_{00} - 1)x_\mu(0) + w_\mu > 0) = \frac{N^+}{N}$,i.e. $P((A_{00} - 1)x_\mu(0) + \sum_{C_j \in N_\mu \setminus C_i} b_j u_j + b_i u_i + z > 0) = \frac{N_{c_i}^+}{N_{c_i}}$ and
- $P((A_{00} - 1)x_\mu(0) + w_\mu > 0 / dropping \quad C_i) = \frac{N_{c_i}^+}{N_{c_i}}$,i.e. $P((A_{00} - 1)x_\mu(0) + \sum_{C_j \in N_\mu \setminus C_i} b_j u_j + z > 0) = \frac{N_{c_i}^+}{N_{c_i}}$

If we consider a random variable $X = (A_{00} - 1)x_\mu(0) + \sum_{C_j \in N_\mu \setminus C_i} b_j u_j + z$, therefore, the probabilities could be expressed as $P(X + b_i u > 0) = \frac{N^+}{N}$ and $P(X > 0) = \frac{N_{c_i}^+}{N_{c_i}}$.
Then, $P(X + b_i u > 0) = P(X > 0) + P(-b_i u < X < 0)$, $P(-b_i u < X < 0) = \frac{N^+ N_{c_i}^- - N^- N_{c_i}^+}{N * N_{c_i}}$, since the output belongs to the closed interval [-1,1] then,

- $P(0 < X + b_i u < b_i u < b_i) = \frac{N^+ N_{c_i}^- - N^- N_{c_i}^+}{N * N_{c_i}}$

The probability of positive output which is bounded by a positive feed-forward parameter corresponding to the i[th] attribute, can be written as $\frac{N^+ N_{c_i}^- - N^- N_{c_i}^+}{N * N_{c_i}}$. From the probability axioms, this term should be greater than zero. Since the denominator is positive, therefore, the nominator should be positive. Accordingly, we were able to prove that the feed-forward parameter behaves inhibitory as a result of $N^- N_{c_i}^+ - N^+ N_{c_i}^-$ being greater than zero. To measure the perturbation happened in the output, it is the percent between the negative and positive outputs. We define a new percent $\alpha = \frac{N^+}{N^-}$ to measure the sign degree and $\alpha_{c_i} = \frac{N_{c_i}^+}{N_{c_i}^-}$ to measure the sign degree by dropping the cell c_i. Then, the sign of the $\alpha - \alpha_{c_i}$ represents the sign measure which can be expressed by dropping the cell C_i the ability to classify more positive rules.

- For example by dropping the cell c_i, in uncoupled CNN, if $\alpha - \alpha_{c_i} > 0$, then more negative rules are classified than positive rules, i.e., the output by dropping the cell c_i is easy to facilitate the negative rules, accordingly, more positive strength is needed. Hence,

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

the feed-forward parameter corresponding to the cell c_i should behave excitatory. For the general case of coupled CNN, as will be explained in the next section, the output is completely determined by $(A_{00}-1)y_{ij}(t_0)+g_{ij}^{+}(t_0)$ and $(A_{00}-1)y_{ij}(t_0)+g_{ij}^{-}(t_0)$, which are similar to the output of uncoupled CNN, hence they follow the same rules.

Since CNN consists of a partial unification of the paradigms Cellular Automata [18] and Neural Network [10], and retaining several elements of both. This new architecture was able to perform time consuming tasks such as image processing and PDE solution, also, it is suitable for VLSI implementation. We can consider the CNN to be a paradigm which is equivalent to Turning Machine. So, it can be completely achieved by constructing the rules that describe its dynamic. To get the minimal decision rule, we have to eliminate the unnecessary conditions in each rule of the algorithm separately.

If ϕ is a C basic formula, $Q \subseteq C$, and ϕ / Q is the Q basic formula obtained from the formula ϕ by removing all the elementary formulas (c_i, u_i^k) such that $c_i \in C-Q$. Then, if $\phi \to \psi$ is a CY decision rule and $c_i \in C$, then c_i is dispensable in $\phi \to \psi$ if and only if $\phi \to \psi$ is satisfied in $Des(C,Y)$ and $\phi \to \psi \in POS(C,Y)$, this implies $\phi / C-\{c_i\} \to \psi$ is also satisfied. Otherwise c_i is indispensable in $\phi \to \psi$. If for all $c_i \in C$ are indispensable in $\phi \to \psi$, then $\phi \to \psi$ will be called independent. So, the subset of attributes $Q \subseteq C$ will be called reduct of $\phi \to \psi$ if $\phi / Q \to \psi$ is independent and $\phi / Q \to \psi$ is satisfied on $Des(C,Y)$, then $\phi / Q \to \psi$ is reduced. As a result of removing the superfluous cells and its corresponding template values, the robustness of the cloning templates should be affected.

- As a result, the algorithm of inducing the optimal structure of Cellular Neural Networks can be demonstrated as follow:

1. Construct the decision table, assuming the problem can be realized by space invariant uncoupled CNN, and calculate the set of possible reduct.
2. if $k=1$, i.e. consistent algorithm then;
   a. Determine the superfluous cells.
   b. Reduce the table according to the reduct set, i.e. remove the attributes that do not belong to the reduct set.
   c. Considering the CNN template structure.
   d. Go to step 4.
   else go to step 3
3. if $k \neq 1$, i.e. inconsistent algorithm, the problem can not be realized by uncoupled CNN, then
   a. Determine the superfluous cells.
   b. Reconstruct the decision table by adding new attributes which represent the output cells that belong to the sphere of influence in the output data in addition to the current reduct of input cells. Exclude the output of the center cell, as it can classify itself.
   c. Check the reduct set again, if $k=1$ go to step 2.
   d. if $k=1-\xi$, ξ is the tolerance, increase the sphere of influence by one and go to step 1.
   else consider a Multi-layer CNN (The future work).
4. Deduce the decision rules that describe the CNN performance.
5. Determine the sign measure and conclude the relation among CNN template parameters.

III. GENETIC ALGORITHMS IN CINSTRAINED OPTIMIZATION

The Induction of the Mathematical System [3], and since some general results have been obtained regarding the effect of the A template on the behavior of the CNN [5], therefore, to guarantee the CNN will converge to a stable equilibrium, it is sufficient to have a sign symmetric A template that is for all $C_{kl} \in N_r(ij)$: $A_{k,l} = A_{-k,-l}$. Also, when $A_{00} > 1$, then all outputs in steady will be either ± 1 and remains in one of the saturations. For robust template we used the randomized search and optimization techniques guided by the principles of evolution and nature genetics, Genetic Algorithms.

Genetic Algorithms, GA, is a stochastic similarities based on sampling techniques especially suited for optimization problem in which a little priori knowledge is available about the function to be optimized. The Genetic Algorithms have be proved to be suitable for complex optimization problems, like combinatorial optimization. In complex optimization problems, an analysis solution is not directly available or a numerical techniques are misled by local minima. The Genetic Algorithms' theoretical foundation lies simply in Darwin's evolutionary explanation of the genesis of species. GA optimization has often guided by blind search; i.e., guided since a reinforcement signal drives it, and blind since it does not access the inside of the signal production itself. Schematically, it works as follows [7] [9]:
A coding is chosen to map any possible candidate solution of a given problem into a finite size string (the chromosome) taken from some alphabet. An initial pool of such string is randomly initialized and each of them is in turn evaluated, ranked according to its capability to solve the given problem. The latter is normally referred to the fitness of the individual and measures what in nature represents an individual's skills in positively interacting with the surrounding environment. The fitness ranking is then used for cloning the genetic material present in the population, i.e. the higher the fitness, the higher the chances that the individual gets its chromosome duplicated and used for mating with other individuals. Mating can be implemented in a variety of ways, but the basic mechanisms are the exchange of sub string in the chromosome (Crossover) and, ,a mutation of the same with a low probability. The newborn individuals then totally or partially replace the old ones in the population, thus a new generation is built. This iterative process is stopped when the maximum fitness in the population does not increase further or has reached a satisfactory value. In either case, the best individual is taken as the solution.

So far, the GA was used for the training of single layer CNN templates [8] [13] [16]. In our research work, we purpose the use of GA for designing CNN [4] structure while asynchronously doing the template learning. Then the results are disjoined and optimized with respect to the robustness. In this case, all the templates are processed on the same external input which is constant during the process, as depicted in Figure 2.

GA codes the candidate problem's solution into a string or *chromosome*. Assuming binary codification, if the maximum number of CNN templates is L, and the number of bits needed to code the template coefficient, m, is related to the range and the precision required. For each CNN template, we defined an additional Boolean parameter, the activation state. If the activation state is set to zero then the corresponding CNN template is deactivated and its template

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

will not be decoded. Thus, the total number of CNN templates of candidate solution will be

$$N_i = \sum_{t=1,\ldots,L} S_t \tag{13}$$

The general form of chromosome can be written as

$$p=[S_1,z^1,b_1^1,\ldots b_9^1,a_1^1,\ldots a_9^1,\ldots,S_L,z^L,b_1^L,\ldots b_9^L,a_1^L,\ldots a_9^L] \tag{14}$$

where the chromosome substring $[z^k, b_1^k,\ldots,b_9^k, a_1^k,\ldots,a_9^k]$ represents the template *k*. The parameters b_i^k, a_i^k are excluded if they are corresponding to superfluous attributes. Since the correct operation of the templates for a given task is achieved by minimizing the error function related to the number of incorrect output pixels, the cost function can be determined by;

$$g(p) = \sum_{i=1}^{k} (y_i^d - y_i(\infty))^2 = \sum_{i=1}^{M}\sum_{j=1}^{N} (y_{ij}^d - y_{ij}(\infty))^2 \tag{15}$$

To achieve the local rules gained by Rough Set, we used the penalty function as new fitness function where the penalty function has the form:

$$\Phi(p) = g(p) + \varphi_1, \tag{16}$$

where $\varphi_1 = \sum_{j=1}^{m} C_j'$, $C_j = (\max\{0, C_j\})^2$, C_j is the inequalities gained by Rough Sets concepts [6]. According to the hardware implementation, the implementation of the CNN-type structure with VLSI chips requires a certain degree of robustness with respect to the mismatching effects. Therefore, the best way to reduce the mismatching effects is by ensuring that the network templates are robust enough. Typically, a relative robustness degree against deviations of the nominal values 5-10% is enough to overcome the mismatch on the VLSI chip. For the definition of the relative fitness of a single layer, we recall the definition in [8]:

$$D(p) = \max_{\alpha}\{\alpha \mid y_\infty(p \circ (1+\alpha 1^{\pm})) = y_\infty(p) \text{ forall} 1^{\pm} \in \beta^j\} \tag{17}$$

where $\circ$ denotes the component wise multiplication, $y_\infty(p)$ is the CNN settle output corresponding to the template p , $\beta = \{-1,1\}$, and $j = (2r+1)(2r+1)+1$. Thus, a total of 2^j possible perturbations of the template set have to be examined for every value of α . In this case, the output is taken once the network had settled to a stable value. To this end, we consider the second cost function as in equation (18) to be a logarithmic distribution because it ensures high penalty to the solutions with robustness under 1% [13], as depicted in Figure 3;

$$g_2(p) = \begin{cases} (1 - \log_{10}(D')) & 0.1\% \le D \le 10\% \\ 0 & D > 10\% \end{cases} \tag{18}$$

$$D' = 100\ D(p)$$

According to the CNN with different number of templates, a linear penalty punishes each solution constrained to the number of templates it codes. If the penalty is excessively strong, significantly better solutions with more templates may be lost, and thus a trade-off between the number of templates and the accuracy of the solution is found, as demonstrated in Figure 4. The general form of the constraint function, can be expressed as follows,

$$g_3(p) = \begin{cases} 0 & \text{if linearly seperable} \\ \frac{N_i}{L} & \textit{otherwise} \end{cases} \tag{19}$$

In order to include these constraints in our algorithm, we will modify the cost function in the way:

$$\tilde{\Phi}(p) = \Phi(p) + g_2(p) + g_3(p) \tag{20}$$

The following features have been used to enforce a more efficient representation phase:

- The best individual from the previous generation substitutes the worst in the current generation if no improvement is made.
- The fitness values are evaluated by equation (20).
- Crossover operator is chosen to be two point crossovers or single point crossover based on the chromosome's length.
- Mutation operator is chosen to be uniform mutation.

IV. EXPERIMENTAL RESULTS

The template learning program has been implemented in Java code. Rough Sets and Genetic Algorithms evaluate every chromosome by discovering the optimal template structure and then computing the transient of the CNN which is defined by the chromosome. Since the computation starts from the same initial state and with the same input values. In the case of a given template, the state equation is integrated every time along the same trajectory in the state space of the network. There are number of parameters in GA which have to be specified. Depending on the application, we can choose our parameters and operators to evolve each generation.

Application 1 (Edge Gray CNN problem)

We decide to apply our method on gray scale input image, where gray scale image contains too much redundancy and required many more "bits" than binary image. For gray-scale input image, the output may not be binary image. Our CNN template called Edge-gray CNN will overcome this problem by accepting gray scale input image and always converging to a binary output image. The Edge-gray problem, depicted in Figure 5, where (a) and (b) refer to the gray-scale input and binary output images respectively. For any gray-scale input image *U*, the corresponding steady state output image *Y* of the Edge-gray CNN, assuming $x_{ij}(0) = 0$, is a binary image, where the black pixels corresponding to pixels laying on the sharp edges of U, or to the fuzzy edges. these edges are defined roughly to be the union of gray pixels of U which form one dimensional (possible short) line segments, or arcs, such that the intensity of pixels on one side of the arc differs significantly from the intensity of neighbor pixels of the other side of the arc.

Experiment was conducted under some conditions; the task was learned with 64 X 64 training example, the population size was 2000 programs, number of generations was 200, and crossover probability, $P_{crossover} = 0.75$, and mutation probability $P_{mutation} = 0.15$. By applying Rough Sets;

1- It is consistent and linearly separable algorithm, so the cloning template can be realized by uncoupled CNN, with 178 different rules, which are able to classify 84 and 94 rules with positive and negative outputs respectively, $\alpha = 84/94 = 0.984$.

2- The reduct set is $\{C_2, C_4, C_5, C_6, C_8\}$, the actual effective cells, with cell significance $\{(1-111/178),(1-114/178),(1-143/178),(1-110/178),(1-115/178)\}$ respectively. Also, the sign measures are as

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

follow $\{(59/52),(54/60),(66/77),(55/55),(56/59)\}$, i.e. {-,-,+,-,-}. So our template is considered as

$$A=\begin{pmatrix}0&0&0\\0&a&0\\0&0&0\end{pmatrix}\quad B=\begin{pmatrix}0&-b_2&0\\-b_4&b_5&-b_6\\0&-b_8&0\end{pmatrix},\ z=R_{real}$$

2.

3- the local rules gained by Rough Sets are summarized by

$$C_5=-1\ \rightarrow y=-1$$

$(C5=1)$, and all the 4 neighbors are black $\rightarrow y=-1$

$(C5=1)$, and at least one of the 4 neighbors is white $\rightarrow y=1$

$(C5\in(-1,1))$, and all the 4 neighbors have the same value as $C5$

$$\rightarrow y=-1$$

Otherwise the output is black.

4- Applying GA on the following problem, we choose the template parameters' intervals related to their significance. Since training data converges to the whole data by a tolerance, then the template parameter should belong to interval with that tolerance. As an example, we choose the [-8,2] interval for negative sign template and [-2,8] for positive sign template. As the result of applying the following to GA, we get the following template

$$A=\begin{pmatrix}0&0&0\\0&2.3&0\\0&0&0\end{pmatrix}\quad B=\begin{pmatrix}0&-1.03&0\\-1.03&4.19&-1.03\\0&-1.03&0\end{pmatrix},\ z=-0.12$$

5- Comparing our method with other previous methods in the literature, such as GA and Truncation learning rules, as demonstrated in Figure 6 below, we found that the Rough Sets increases the GA convergence, as an expected result for reduction of the number of parameters. Also combining both of Rough Sets and GA improves the fitness function, as a result of increasing the robustness of our template. The comparison among different techniques is declared in Table 1. We defined the comparison criterion as the percent of error occurred as the result of the robustness changes on the template parameters and the number of iteration that are needed for each cell to enter the saturation region. Also, we extended the comparison to handle the ability to discover the optimal template structure. As a result of our comparison, we are able to say that GA always needs other methods to complete its shortcoming. Also, the truncation learning rules perform the same as GA.

Application 2 (Shadow detection CNN)

An example for propagating type templates is the shadow detector [14], in each row, all the pixels right from the left most black pixels should become black. The training set is shown in Figure 7, where (a) refers to the input image, and (b) is the desired output.

Experiment was conducted under some conditions; the task was learned with 20×20 training example, the population size was 500 chromosomes, number of generations was 100, crossover probability $P_{crossover}=90\%$, and mutation probability was $P_{mutation}=0.01$. By applying Rough Sets we get the following;

We found that, by applying Roughs Set concept on the decision table, it's in inconsistent algorithm and it has $k=0.6$ of consistent rules. We discovered Completing the decision table by adding the output corresponding to the reduct set, i.e. the new attributes became $C=\{c_0,c_1,...c_8,y_1,...y_4,y_6,...,y_8,y_9\}$, where the output of the cell itself is removed, and then checking the consistency of the modified table.

Then, it is a consistent algorithm, $k=1$, with four true rules, three positive and one negative rules respectively, the reduct set is, $\{y_4,y_5\}$. Also, the sign measures are as follow $\{(2/0),(2/0)\}$ which indicate that both of them will behave similarly. Measuring the stability indicates that the self feedback should be positive and greater than one, both template parameters are considered to be positive. So our template is considered as

that the cell C_9 is superfluous cell.

$$A=\begin{pmatrix}0&0&0\\A_4&A_5&0\\0&0&0\end{pmatrix}\quad B=\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix},\ z=R_{real}$$

3. The dynamic rules gained by Rough Sets are summarized by:

If (the input cell C5 is white and its neighbor output cell C4 is white) implies the output is white.

If (the input cell C5 is white and its neighbor output cell C4 is Black) implies the output is Black.

If (the input cell C5 is Black) implies the output is Black.

4. Applying GA, the template parameter are generated as below with robustness 35%;

$$A=\begin{pmatrix}0&0&0\\3.59&4.654&0\\0&0&0\end{pmatrix}\quad B=\begin{pmatrix}0&0&0\\0&0&0\\0&0&0\end{pmatrix},\ z=6.868$$

Application 3 (Image Enhancement)

According to the noisy acquisition devices and variation in impression conditions, the ridgelines of fingerprint images are mostly corrupted by various kinds of noise causing cracks, scratches and bridges in the ridges as well as blurs. This application is to demonstrate the ability of our method to enhance the grey scale finger print images by removing the undesired noises.

Since the input in this case is a grey scale pattern, it is impossible to take into account the all possible inputs combinations when calculating the robust templates. The approach used here consists of considering only the possible input values contained in the training pattern. Thus, the training patterns must be carefully selected not only to define the task under consideration, but also to contain relevant information about the patterns to be processed. A fingerprint pattern of size 592 * 614 is selected, as illustrated in Figure 8. In Figure 8, the input is shown in Figure 8 (a) and the enhanced image is shown in Figure 8 (b).

1. By applying Rough Set concepts, we get inconsistent algorithm with 127106 different true rules, K= 0.84022 with no superfluous cells.
2. By expanding the decision table to include the neighboring output pixels as classified attributes, we get inconsistent Algorithm with 176457 different true rules, K= 0.99176. Thus a single layer can't realize the desired goal.
3. The GA is used to discover the first layer. By running our experiment under constrained optimization we get the following templates

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$A = \begin{bmatrix} 1 & 6 & 7 \\ 2 & 3 & 1 \\ 1 & 1 & 2 \end{bmatrix} B = \begin{bmatrix} 1 & 1 & 5 \\ -4 & -1 & 3 \\ -5 & 2 & 1 \end{bmatrix}, z = -1$$

4. By Applying Rough Set to conclude the number of different layers remain to enhance the image, we get consistent algorithms with the following templates.

$$A = \begin{bmatrix} 1.02 & 3.22 & 4.62 \\ 2.82 & 1 & 3.68 \\ 2 & 2.88 & 5.58 \end{bmatrix} B = \begin{bmatrix} 3.5 & 2.73 & 3.96 \\ -4.94 & -5.43 & 2.11 \\ 4.83 & -3.99 & 1.57 \end{bmatrix}, z = -1.47$$

Application 4 (Image Half-toning)

Half-toning [11] is the process of coding gray-scale images by the binary (black-white) value at each pixel. Upon display, it is required that, by the blurring of the eye, the half-tone image will appear similar to the original continuous toned image. This process is required in many applications where the displayed medium can only support binary output. For instance, photographic half-toning techniques have long been used in newspaper printing where the resulting binary values represent the presence or absence of black ink. Digital image halftones are required in many present day electronic applications such as FAX (facsimile), electronic scanner/coping, laser printing and low band width remote sensing. This application is to demonstrate the ability of our method to recognize a propagating type template. According to our method, this template can not be recognized by a single layer with 3×3 but it can be recognized by 5×5 as shown below;

1. At the first stage, by applying Roughs Set concept on the decision table, it is inconsistent algorithm with $k = 0.876$ of consistent rules without superfluous cell. The reduct set is equal to $\{C_0, C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9\}$.
2. Completing the decision table by adding the output corresponding to the reduct set, i.e. the new attributes became $C = \{C_0, C_1, ...C_8, C_9, y_1, ...y_4, y_6, ..., y_8, y_9\}$ where the output of the cell itself is removed, and checking the consistency of the modified table.
3. An inconsistent algorithm has been discovered, $k = 0.951$, with 706 different rules, 362 positive rules and 344 negative rules. The reduct set is given by $\{C_0, C_1, C_3, C_7, C_9, y_1, y_2, y_3, y_4, y_6, y_7, y_8, y_9\}$ with sign measure
4. {321 318 323 319 322 303 306 307 310 301 304 300 /312 /308 /314 /310 /310 /287 /288 /281 /284 /286 /280 /285}, i.e.{+,+,+,+,+,-,-,-,-,-,-,-,-}. Then the optimal template structure, when the initial state is considered as the image itself, to recognize 95% of the correct output is considered as

$$A = \begin{pmatrix} -A_1 & -A_2 & -A_3 \\ -A_4 & A_5 & -A_6 \\ -A_7 & -A_8 & -A_9 \end{pmatrix} \quad B = \begin{pmatrix} b_1 & 0 & b_3 \\ 0 & b_5 & 0 \\ b_7 & 0 & b_9 \end{pmatrix}, z = R_{real}$$

By Applying GA, with the population size $pop = 6000$, the number of generation 300, crossover probability $P_{crossover} = 0.8$ and mutation probability $P_{mutation} = 0.1$, we get the following template where the similarity among the template parameters is considered;

$$A = \begin{pmatrix} -0.38 & -0.38 & -0.38 \\ -0.38 & 1.209 & -0.38 \\ -0.38 & -0.38 & -0.38 \end{pmatrix} \quad B = \begin{pmatrix} 0.44 & 0 & 0.44 \\ 0 & 2.54 & 0 \\ 0.44 & 0 & 0.44 \end{pmatrix}, z = -0.11$$

5. By expanding the radius of the influence sphere r to be two and by applying Rough Sets concepts, it is inconsistent algorithm with degree of dependencies $k = 0.936$ of consistent rules without superfluous cell. The reduct set is equal to $\{C_0, C_1, C_2, C_3, ..., C_{23}, C_{24}, C_{25}\}$.

6. Completing the decision table by adding the output corresponding to the reduct set, i.e. the new attributes became $C = \{C_0, C_1, ..C_{24}, C_{25}, y_1, y_2 .. y_{12}, y_{14}, ..., y_{23}, y_{24}, y_{25}\}$ where the output of the cell itself is removed, and checking the consistency of the modified table.

7. It's consistent algorithm, degree of dependencies $k = 1$, with the following structure

$$B = \begin{pmatrix} b1 & 0 & b3 & 0 & b5 \\ 0 & b7 & 0 & b9 & 0 \\ b11 & 0 & b13 & 0 & b15 \\ 0 & b17 & 0 & b19 & 0 \\ b21 & 0 & b23 & 0 & b25 \end{pmatrix} \quad A = \begin{pmatrix} -a1 & -a2 & -a3 & -a4 & -a5 \\ -a6 & -a7 & -a8 & -a9 & -a10 \\ -a11 & -a12 & a13 & -a14 & -a15 \\ -a16 & -a17 & -a18 & -a19 & -a20 \\ -a21 & -a22 & -a23 & -a24 & -a25 \end{pmatrix} z = R_{real}$$

8. Applying GA with considering similarity relation, we get the following templates;

$$B = \begin{pmatrix} 0.125 & 0 & 0.49 & 0 & 0.125 \\ 0 & 0.395 & 0 & 0.395 & 0 \\ 0.49 & 0 & 2.65 & 0 & 0.49 \\ 0 & 0.395 & 0 & 0.395 & 0 \\ 0.125 & 0 & 0.49 & 0 & 0.125 \end{pmatrix}$$

$$A = \begin{pmatrix} -0.069 & -0.112 & -0.129 & -0.112 & -0.069 \\ -0.112 & -0.296 & -0.556 & -0.296 & -0.112 \\ -0.129 & -0.556 & 1.20 & -0.556 & -0.129 \\ -0.112 & -0.296 & -0.556 & -0.296 & -0.112 \\ -0.069 & -0.112 & -0.129 & -0.112 & -0.069 \end{pmatrix} \quad z = -0.05$$

V. CONCLUSION

In this paper, a new learning method for discovering the optimal CNN templates is proposed. Rough Sets and Genetic Algorithms are integrated in learning the CNN template to overcome the shortcoming caused by each of them. The idea is to describe the CNN dynamic by a decision table and then to use the concept of Rough Sets in deducing the optimal CNN structure. This is achieved by removing the superfluous cells that have no affect on classifying the output, based on determining the significance of each cell. Our algorithm relies on discovering the consistency relation among the rules, by means of decision language, and then determining the dependencies among data. The reduced decision rules, decision algorithm, that specify the space invariant CNN dynamic are derived. Also, the reduced decision rules are used in discovering which algorithm can be realized by uncoupled CNN or by coupled CNN based on modifying the decision table by adding new attributes to evaluate the optimal CNN structure with propagating type. Since the new method relies on modifying the decision table by adding new attributes, i.e. the new attributes are considered in the saturation regions and away from the linear region. A new measure, the sign measure, has been introduced to demonstrate the relation among the template

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

parameters. Depending on the local rules that are discovered by Rough Sets, the comparison principle technique is used in discovering an affine system of inequalities. This system of inequalities must be satisfied by the parameters of the templates to ensure a correct operation of the CNN. Because of the sensitivity of the templates to small variation around their nominal value, GA with constrained fitness function is used in learning the templates in propose of yielding more robust template. The GA could generate simple template, but the number of free parameters in the template increase its performance break down, therefore, the GA chromosome structure is chosen in accordance with the number of affective attributes. Also, the GA parameters' ranges are considered in accordance with the sign measure. The chromosomes were evaluated according to the transient behaviour of the CNN and the Performance of the chromosome is determined by a penalty fitness function. It is determined by means of the quadratic difference between the desired output and the settled output of the CNN in addition to, constraints on the system of inequalities and the robustness issues. The new method is applied on four different application problems, Edge Gray CNN, Shadowing, image enhancement and Image Half-toning. The result of the new introduced method provides the ability of discovering the solution for a problem of any domain. Moreover, the compression between the new method and other previous methods, such as GA and Truncation Learning algorithms, is declared to demonstrate the efficiency of the new introduced method. Possible extension of the proposed method is to improve the templates with only integer values by means of integer programming algorithm. This is very advantageous from a chip designer perspective, where for the programmability of CNN hardware is usually not continuous but restricted to a discrete set of values, namely the integers and a few simple rational numbers. Also, we consider presenting a general framework to handle the general problem of multi-layer CNN in our future work.

REFERENCES

[1] Chua, L. O. "CNN: A vision of complexity", International Journal of Bifurcation and Chaos, vol. 7, no. 10, 2219-2425, 1997.

[2] Chua, L. O. and L. Yang, "Cellular Neural Networks: Theory and applications", IEEE Transaction on Circuits and System. vol. 35, no. 10, pp. 1257-1272, Oct., 1988.

[3] Chua, L. O. and Patrick Thiran," An Analytic Method for Designing Simple Cellular Neural Networks", IEEE Transaction on Circuit and Systems, vol. 38, no. 11, pp. 1332-1341, 1991.

[4] Chua, L. O. and Tamas Roska,." Cellular Neural Networks and Visual Computing", Cambridge University Press, 2002.

[5] Civalleri, P.P. and M. Gilli," On stability of Cellular Neural Networks", Journal of VLSI Signal Processing, vol. 23, pp. 429-435, 1999.

[6] Elsayed Radwan and Omaima Nomir, "An Analytical Method for Learning Cellular Neural Networks based on Rough Sets", Proceeding of ICCTA 2007, Alexandria, Egypt, pp. 19-22, 2007

[7] Goldberg, D. E. "Real –coded Genetic Algorithms", virtual alphabets and blocking", Complex Systems, vol. 5, 139-167, 1991.

[8] Hanggi, M., Moschytz G.," Cellular Neural Networks: Analysis, Design, and Optimization", Kluwer Academic Publishers: Dordrecht, MA, 2000.

[9] Holland, J. H., "Adaptation in Natural and Artificial Systems (1992 edition)", Cambridge, MI: MIT Press, 1992.

[10] Hopfield, J. J., "Neural Networks and Physical Systems with Emergent Computational Capabilities", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 79, pp.2554–2558, 1982.

[11] Kenneth R. Crounse, Tamas Roska and Leon O. Chua, " Image Halftoning with Cellular Neural Networks", IEEE Transactions on Circuits and Systems-II: Analog and Digital Signal Processing, vol. 40, no. 4, pp. 276-283, 1993

[12] Lech Polkowski, Shusaku Tsumoto, TsauY. Lin," Rough Sets Methods and Applications: New Development in Knowledge Discovery in Information Systems", Physica-Verlag Heidelberg, 2000.

[13] Lopez, P., D.L. Vilarino, V. M. Brea and D. Cabello," Robustness Oriented Design Tool for Multi-Layer DTCNN Applications", International Journal of Circuit Theory and Applications, vol. 30, pp. 195-210, 2002.

[14] Matsumoto, T., L. O. Chua, and H. Suzuki, "CNN cloning template: shadow detector", Transaction on Circuits and Systems, vol. 37, pp. 1070-1073, 1990.

[15] Pawlak, Z. "Rough Sets Theoretical Aspects of Reasoning about Data", Kluwer Academic Publishers, 1991.

[16] Tibor Kozek Tamas Roska, and Leon O. Chua, Genetic Algorithms for CNN Template learning", IEEE Transactions on Circuits and Systems, vol. 40, no.6, 392-402, Jun. 1993.

[17] Winter, G., J. Periaux, M. Galan and P. Cuesta, "Genetic Algorithms in Engineering and Computer Science", John Wiley & Sons Ltd., 1995.

[18] Wolfram, S.," Cellular Automata as Models of Complexity", Nature, vol. 311, pp. 419-424, October 4, 1984.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Fuzzy-Based Dynamic Rough Set Resource Discovery According to User Preferences in Grid Environment

Iraj Ataollahi
Computer Engineering Department
Iran University of Science and Technology
Tehran, Iran
ir_ataollahi@mail.iust.ac.ir

Mahdi Bakhshi
Department of Computer
Islamic Azad University, Shahrbabak Branch
Shahrbabak, Iran
mah.bakhshi@yahoo.com

Morteza Analoui
Computer Engineering Department
Iran University of Science and Technology
Tehran, Iran
analoui@iust.ac.ir

***Abstract*—Grid environment is a service oriented infrastructure in which many heterogeneous resources participate to provide high performance computation. One of the bug issues in the grid environment is the uncertainty among registered resources. Furthermore, in an environment such as grid dynamicity is considered as a crucial issue which must be dealt with. Using dynamic rough set theory to deal with uncertainty and dynamicity has shown good results. Addition to this, compounding this theory with Fuzzy system will improve its efficiency by applying fuzzy system in matchmaking phase. In this work we propose a solution, called Fuzzy-Based Dynamic Rough Set Resource Discovery (FDRSRD), in which dynamic Rough set theory is used in order to deal with uncertainty and fuzzy system for resource matchmaking according to the user preferences. In matchmaking phase, in order to improve accuracy and speedup, fuzzy system is used to rank resources. We also report the result of the solution obtained from the simulation in GridSim simulator along with FuzzyJ Toolkit. The comparison has been made between FDRSRD, Dynamic and Classical Rough Set based algorithms. FDRSRD shows much better precision and more speed for the cases with uncertainty in a dynamic system such as the grid rather than the two other algorithms.**

Keywords- Grid; Rough Set; Dynamic Rough Set; Resource Discovery; Ontology; Fuzzy System; User Preferences

I. INTRODUCTION

Nowadays, Grid is considered as a service-oriented computing infrastructure [1]. Open Grid Services Architecture (OGSA) [2] has been used for dealing with service-oriented problem [3]. OGSA has been improved by Global Grid Forum.

Many resources such as workstations, clusters, and mainframes with various properties such as main memory, CPU speed, bandwidth, virtual memory, hard disk, cost, and response time etc are joining and leaving the grid environment. On the Other hand many users want to use these resources to run their jobs with different requirements. But there are always differences between which a user requested and which have been registered in the Grid Information Server (GIS). This may lead to uncertainty in GIS data base which will be resulted in to the reduction of the precision of the matchmaking algorithms. To solve this vagueness and uncertainty we use rough set theory, proposed by Z. Pawlak in 1982 [4], which has been used in vast area of computer science such as data mining, pattern recognition, machine learning and knowledge acquisition etc [5].

One of the first methods that can be used for service discovery is UDDI which is used for web service publication and discovery. The current web service discovery mechanism is based on the standard of UDDI [6]. In UDDI, XML is used to describe data in business services. UDDI process searches queries according to keywords and classification information. There is limitation with the discovery mechanism of UDDI. Firstly, machine can read XML data, but it can not understand XML data. Different query keywords may be semantically equivalent, whereas UDDI can not infer any information from keywords or tModels it can easily make mistake. Secondly, search by keywords and taxonomy is not suitable for web service discovery. Furthermore, UDDI does not support search by service capabilities and other properties [7]. This makes UDDI search method a low precision method [6].

By advent of semantic web, services can be annotated with metadata for enhancement of service discovery. One of the earliest to add semantic information is DAML-S [8]. DAML-S uses semantic information for discovering Web services. DAML-S uses ontological description to express web service capacity and character.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

OWL-S is an OWL [9] based ontology for encoding properties of Web services. OWL-S technology is used to facilitate service annotation and matching. OWL-S ontology defines a service profile for encoding a service description, a service model for specifying the behavior of a service, and service grounding for how to invoke the service. Actually, by using domain ontology described in OWL, using special software such as protégé [10], a service discovery process involves a matching between the profile of a service advertisement and the profile of a service request. The service profile describes the functional properties such as inputs, outputs, preconditions, and effects, and non functional properties such as service name, service category, and aspects related to the quality of service.

In [11] a quantification standard for semantic service matching has been presented that modifies the classical matching algorithm based on OWL-S. Matching algorithm has used the quantification standard of service matching and OWL-WS. In [12] service composition algorithm has constructed a mathematical model and converted it to the shortest path problem in order to find process that can satisfy customer need in best conditions.

In [7] an approach has been developed for integrating semantic features into UDDI. The approach uses a semantic matchmaker that imports OWL-S based semantic markups and service properties into UDDI. The combination of OWL-S and UDDI shows there could be a service discovery which supports web service expression while UDDI is used. The matchmaker, therefore, enables UDDI to store semantic information of web services and process service search queries based on semantic similarity of web service properties [7].

The above-mentioned methods facilitate service discovery in some way. However, when matching service advertisements with service requests, these methods assume that service advertisements and service requests use consistent properties to describe relevant services. But for a system such as Grid with a large number of resources and users which have their own predefined properties to describe services, it can't be true that service advertisements and service requests use consistent properties to describe services. In other words, some properties may be used in service advertisement that may not be used by service request. So, an approach must be taken into consideration to deal with uncertainty of service properties when matching service advertisements with service requests.

Rough set theory is a new mathematical theory which deals with uncertainty and vagueness [13].

By moving toward the age of information, a hypothesis can formulate the human knowledge in the systematic form, and introduce an approximate description that is reliable and analyzable. This important subject is applicable by a fuzzy system [14]. In our previous works [15, 16], we have used classic and dynamic rough set theory to deal with uncertainty and vagueness. Using fuzzy system in matchmaking phase make it possible to find resources met the requested properties with more precision and speed. Matchmaking phase is done by using fuzzy system expressed according to properties of registered resources and user preferences which assigned to these properties.

The remainder of this paper is organized as fallows. Part II is a review of related works. Part III is a description of the classical rough set theory, part IV is a description of the dynamic rough set theory. Part V is the description of algorithm implemented and used in this paper, part VI is a comparison of our algorithm with dynamic and classical rough set based matchmaking algorithms [15,16], and finally part VII is the conclusion and future works.

II. Related Works

While the grid environment moves towards a service-oriented computing infrastructure, service discovery is becoming a vital part of this environment. One of the earliest methods for service publication and discovery is UDDI which only supports keyword matches and does not support any semantic service. DAML-S is the earliest to add semantic information for discovering web services [18]. DAML-S offers enough semantic information expressing Web service capacity and character with ontological description of web services. In past few years, a great amount of studies have been carried out on the basis of OWL-S, such as semantic expression service bundling [19], ontology-based service matching [19], OWL-S and UDDI combination [17]. In the [21] a metric is proposed to measure the similarity of semantic services annotated with OWL ontology. Similarity is calculated by defining the intrinsic information value of a service description based on the inferencibility of each of OWL constructs. All the above methods do not support uncertainty in properties. Rough set theory is used for dealing with vagueness and missing data in large variety of domains. So, compared with the work mentioned above, rough set theory can tolerate uncertain properties in matching resources. In the [16] classical rough set based algorithm has been proposed to deal with uncertainty and vagueness. Whereas grid environment is dynamic, using classical (static) rough set theory can not seem proper to use [15]. In this paper, our algorithm works in two steps. The First step is candidate optimization component which optimize the candidate set using dynamic rough set theory. The Second step is fuzzy matchmaking component which ranks resources using fuzzy system according to requested resource.

III. Classical Rough Set Theory

Rough set theory which is proposed by Pawlak, in 1982, has been proved to be a good mathematical tool to describe and model uncertainty and imprecision. It has been widely applied in artificial intelligent, pattern recognition, data mining, fault diagnostics etc [21]. There are many advantages of rough sets theory; for example, no preliminary or additional information is needed and only the facts in the data are considered.

Let:

- U: a set of N registered resources, U= $\{u_1, u_2, \ldots, u_N\}$, $N \geq 1$.
- P: a set of M properties used to describe the N registered resources of the set U, P = $\{p_1, p_2, \ldots, p_M\}$, $M \geq 2$.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

- Q: a set of K registered resource properties relevant to a resource request R in terms of resource ontology whose irrelevant properties have been removed, $Q = \{q_1, q_2, \ldots, q_K\}$, $K \geq 1$, and Q is a subset of P.
- R: a set of L requested resource properties with their weights, $R=\{(r_1,w_1), (r_2,w_2), \ldots, (r_L,w_L)\}$, $L \geq 1$.

According to the rough set theory for a given set X there are:

$$\underline{Q}X = \{x \mid [X]_Q \subseteq X\} \qquad (1)$$

$$\overline{Q}X = \{x \mid [X]_Q \cap X \neq \phi\} \qquad (2)$$

In which $\underline{Q}X$ is the lower approximation and $\overline{Q}X$ is the upper approximation of X in terms of properties set Q. X is a subset of U and Q is a sub set of P.

$X \subseteq U$

$Q \subseteq P$

So for a property $q \in Q$, we can say that:

- $\forall x \in \underline{Q}X$, x definitely is a member of X and definitely has property q.
- $\forall x \in \overline{Q}X$, x probably is member of X and probably has property q.

$\forall x \in U - \overline{X}$, x absolutely is not a member of X and absolutely does not have property q.

The Most important part of rough set theory is attribute reduction. Some attributes are dependent on other attributes in attributes set, so they are not necessary to be considered in matching phase. According to rough set theory we are:

$$POS_C(D) = \bigcup_{X \in U/D} \underline{C}X \qquad (3)$$

$$\alpha = \gamma(C,D) = \frac{|\underline{C}X|}{|U|} \qquad (4)$$

In Which C and D are subsets of property set P. as shown in [13], D totally depends on C if α =1 Or D partially (in a degree of α) depends on C if $\alpha < 1$.

Since existing works need to find exact match between requested resources and registered resources, it is difficult to find exact matching. So by using rough set theory, the need of exact match has been removed.

IV. Dynamic Rough set Theory

Although rough set theory is being used in various ranges of research such as data mining, pattern recognition, decision making and expert system, it is suitable for static knowledge and data. In fact, in a classical rough set theory, subset X of universal set U is a static set without considering the dynamic properties it can have. In the real word, most information systems have dynamic properties so that the rate of participant and disappearance of entities in these systems is high. Whereas Pawlak's rough set theory can only deal with static information system, using a dynamic method to deal with uncertainty and process information system will have more efficiency.

By using dynamic rough set theory, considering dynamic properties of an information system will be possible. Dynamic rough set theory uses outward and inward transfer parameters to expand or contract X set in classical rough set.

According to [22], dynamic rough set theory has been defined as follows:

Suppose A= (U, P) is an information system, $T \subseteq P$ and $X \subseteq U$. For any $x \in U$, we have:

$$\rho^{-}_{(X,T)}(x) = \frac{|[x]_T - X|}{|[x]_T|}, \text{as } x \in X \qquad (5)$$

$$\rho^{+}_{(X,T)}(x) = 1 - \frac{|[x]_T - X|}{|[x]_T|}, \text{as } x \in \sim X \qquad (6)$$

$\rho^{-}_{(X,T)}(x)$ is called outward transfer coefficient and $\rho^{+}_{(X,T)}(x)$ is called inward transfer coefficient of element x about T. In real computation, outward and inward transfer coefficients are been choose as constant amounts. In fact $d^{-}_{T}(X) \in [0,1]$ and $d^{+}_{T}(X) \in [0,1]$ are outward transfer standard and inward transfer standard of elements of X about T, respectively.

Inflated dynamic main set of X is defined as below:

$$M^{+}_{T}(X) = \{x \mid x \in (\sim X), d^{+}_{T}(X) \leq \rho^{+}_{(X,T)} < 1\}. \qquad (7)$$

And inflated dynamic assistant set is defined as:

$$A^{+}_{T}(X) = \{x \mid x \in (\sim X), 0 \leq \rho^{+}_{(X,T)} < d^{+}_{T}(X)\}. \qquad (8)$$

X^{+}_{T} is called inflated dynamic set of X about T and defined as:

$$X^{+}_{T} = X \cup M^{+}_{T}(X). \qquad (9)$$

The formulas (5-9) show that we can expand X according to T. we can also contract X according to T. for this reason we have:

$$M^{-}_{T}(X) = \{x \mid x \in X, d^{-}_{T}(X) \leq \rho^{-}_{(X,T)}(X) < 1\}. \qquad (10)$$

In which $M^{-}_{T}(X)$ is defined as contracted dynamic set of X about T and also contracted dynamic assistant set is defined as:

$$A^{-}_{T}(X) = \{x \mid x \in X, 0 \leq \rho^{-}_{(X,T)}(X) < d^{-}_{T}(X)\}. \qquad (11)$$

And X^{-}_{T} called contracted dynamic set is defined as:

$$X^{-}_{T} = X - M^{-}_{T}. \qquad (12)$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

According to the above mentioned, we can expand and contract X according to T. Suppose we have T and $T' \subseteq P$, two direction dynamic set of X according to the T and T' is defined:

$$X^*_{(T,T')} = (X - M^-_T(X)) \bigcup M^+_T(X). \qquad (13)$$

Suppose $Q \subseteq P$, we can compute upper and lower approximation of $X^*_{(T,T')}$ using equations (1, 2) so that we have:

$$\underline{Q^*_{(T,T')}}(X) = \{x \mid x \in U, [x]_Q \subseteq X^*_{(T,T')}\}. \qquad (14)$$

$$\overline{Q^*_{(T,T')}}(X) = \{x \mid x \in U, [x]_Q \cap X^*_{(T,T')}\} \qquad (15)$$

$\underline{Q^*_{(T,T')}}(X)$ and $\overline{Q^*_{(T,T')}}(X)$ are called two direction transfer D-lower approximation set and two direction transfer D-upper approximation set of X, respectively.

In fact according to $M^+_T(X)$ we should increase resources (X) which can have opportunity of selection according to the attributes set T, but $M^-_{T'}(X)$ indicates according to the attributes set T' we should decrease X.

the $\underline{Q^*_{(T,T')}}(X)$ indicates the objects of the optimization of the candidate set which can be considered as a candidate set for matchmaking process. So in the matchmaking phase we only need to search D-lower approximation set ($\underline{Q^*_{(T,T')}}(X)$) in order to select resources which satisfy requested service.

In this work, we can also determine the priority of each requested service property, so that if properties T have an important role, their priority factor is high, we can decrease d^+_T and this means that we expand candidate set X according to the properties set T. when T' plays a less important role, priority of properties is low, we can decrease $d^-_{T'}$ in order to contract the candidate set.

V. Resource Discovery

GridSim simulator has been used in order to simulate Dynamic Rough Set Resource Discovery Algorithm (DRSRD). As shown in Fig. 1, user sends a service request to the GridSim's Broker, Broker forwards the request to the GIS which can access Advertised Resource Repository and Ontology template in order to get resources which satisfy requested service. GIS has two components in order to find resources satisfying requested service. First component is Candidates Optimization which uses dynamic rough set theory in order to determine the optimum set of candidate resources. User defines a priority factor called W_i for each of the requested service properties in order to determine their priority. Candidate optimization component determines candidate resources set according to the priority of requested service properties.

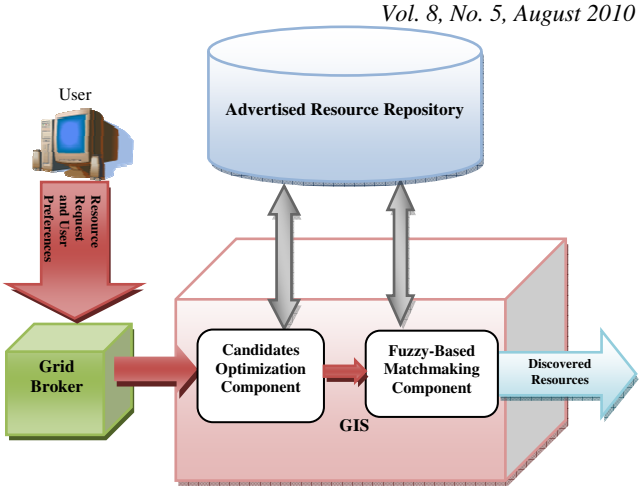


Figure 1. Algorithm outline

The Second component is the Matchmaking component which does the matchmaking algorithm using fuzzy system on the candidate resources set obtained from the candidates optimization component.

For describing resource properties, we have used a resource ontology template based on the Karlsruhe ontology model [10]. The resource ontology template has been created by considering the most possible computing resources in the Grid. The concept of these resources has been defined properly using relations and properties so that the characteristics of any resource can be defined by their properties. For using the ontology template in the GridSim, which is a java base simulator, we have used the protégé-OWL API, which is a java base API, in order to create and modify Ontology dynamically.

In this section we will describe the candidate optimization component and matchmaking component.

A. Candidates Optimization

The Most important aim of dynamic rough set theory is to deal with the vagueness and uncertainty in a knowledge system which changes dynamically. For a system such as the Grid whose resources can join or leave the system randomly, using dynamic rough set theory is more efficient than classical rough set theory.

User sends its service request to the Broker. In this request, each one of the requested service properties has a weight Wi. Broker forwards this request to the Grid Information Service (GIS) in order to find the best resources which satisfy the requested service. After getting the request by GIS, it classifies the requested properties according to their weight. According to part III, the set R is the requested resource properties and the properties set T which $T \subseteq R$ is defined as bellow:

$$T = \{(r_i, w_i) \mid (r_i, w_i) \in R \text{ and } w_i \geq 0.5\},\ 1 \leq i \leq L1\}$$

In fact the set T contains properties with priority factor (weight) more than 0.5.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

As mentioned in part IV the candidate set can be expanded according to the properties set T. According to the weight of requested service properties, we define the inward transfer standard $d_T^+(X)$ as follows:

$$d_T^+(X) = \frac{\sum_{i=1}^{L1} w_i}{|T|}, which\ (t_i, w_i) \in T \qquad (16)$$

The properties set T', in which $T' \subseteq R$, are defined as a set of properties the weight of which is less than 0.5. So T' is defined as:

$$T' = \{(r_i, w_i) | (r_i, w_i) \in R \text{ and } w_i < 0.5\},\ 1 \le i \le L2.$$

The outward transfer standard $d_{T'}^-(X)$ is defined as bellow:

$$d_{T'}^-(X) = \frac{\sum_{i=1}^{L2} w_i}{|T'|}, which\ (t_i, w_i) \in T' \qquad (17)$$

The candidates set X is defined as a set of resources with maximum non empty properties according to the requested resource properties. And ~X is defined as all resources in the universal set U which are not contained in the X.

Candidates Optimization algorithm is shown in the Fig. 2. Algorithm uses three steps to compute candidates optimized set.

```
Input: requested properties set R={(r1,w1), (r2,w2), …,(rL,wL)}.
Input: candidates set X.
Output: candidates optimized set.
  I: Inflated dynamic main set of X about T .
  C: contracted dynamic set of X about T' .
  X* : Two direction dynamic set of X according to the T and T' .
  _X*_ : Lower approximation of X* according to requested resource properties R.

  Step 1:
     Compute d+T(X) and d-T'(X) .
  Step 2:
     For all x∈~X
        If ρ+(X,T)(x) ≥ d+T(X)
           Add x to the I.
     End for.
     For all x∈X
        If ρ-(X,T')(x) ≥ d-T'(X)
           Add x to the C.
     End for.
  Step 3:
     X* = (X − C) ∪ I .
  Step 4:
     Compute _X*_ according to the R.
  Return _X*_ .
```

Figure 2. Candidates Optimization algorithm

Step 1 calculates $d_T^+(X)$ and $d_{T'}^-(X)$ using the equations (1) and (16) respectively. In step 2, the inflated dynamic main set of X and contracted dynamic main set of X using equations (7) and (10) respectively.

Step 3 calculates two direction dynamic set of X according to T and T' using equation (13). Candidates set X can be expanded according to the properties set T which has properties with higher priority and can be contracted according to the properties set T' the properties of which have lower priority. In Step 4, by using equation (14), the lower approximation set $\underline{X^*}$ of X^* is calculated according to the requested resource properties set R. In fact $\underline{X^*}$ is a set of resources that are most likely to be selected as matched resources.

B. Resource Matchmaking

After optimization of the candidates set we should only apply the matchmaking algorithm on the optimized candidates set. Reduction of the candidates set causes the reduction of searching time.

We design matchmaking algorithm according to the fuzzy logic in which resource ranking is done using a fuzzy system which fuzzy rules have been expressed according to user preferences.

1) Definition of the Variables and Membership Functions of the System

Vague knowledge, i.e. rules based on fuzzy logic, are also important from the perspective of evaluating values of attributes that have very complex dependencies with other attribute values. The vague membership functions can be modeled in the form of some sets by fuzzy logic [23].

On the other hand, in simplest form, a domain ontology would specifies the valid vocabulary of describing (naming) functional and nonfunctional properties that are allowed to occur in resource descriptions, but we need a domain ontology that can help in defining categories through linguistic variables. For example, the response time could be described with the terms *fast, normal, slow, very slow* [24].

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

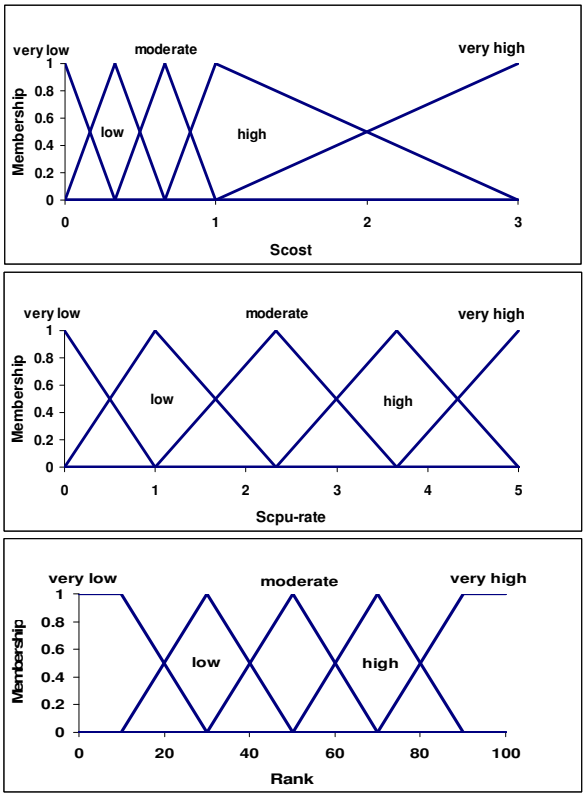


Figure 3. membership functions for defining linguistic variables of the system

With complete knowledge of linguistic variables, we can define the membership functions of these variables.

In matchmaking component in order to correct operation of fuzzy system we define domain of membership functions which are related to the properties of registered resources in the form of the proportion defined as:

$$S_i = \frac{P_i}{R_i} \qquad (18)$$

In which P_i and R_i are the value of the registered and requested property i, respectively. S_{cost} and $S_{cpu\text{-}rate}$ are examples of input variables and rank is example of output variable, as shown in Fig.3.

To define membership functions in this approach, it is important to use equal terms for definition of system's linguistic variables. It is considerable, because of logical relationship between the input and output variable terms in the fuzzy rules formation of the system.

2) Modeling User Preferences Based on Weighting the Rules

We view preferences as the information that describes the constraints on the properties of an individual in order to be accepted for further consideration. We specify different levels of acceptance with fuzzy membership functions.

This approach do resource matchmaking with assumption of the existence of the fuzzy rules that can be criteria for ranking properties related to resources, gives more weight to the rules that are more important from user's point of view.

The confidence factor (CF) of every rule which is a number between 0 and 1, can express the confidence value and the importance of the rule to obtain the final result. Equation (19) expresses the effect of this factor in computing the result [25].

$$Membership_{con,i} = Membership_{premise,i} \times CF_i \qquad (19)$$

This equation shows that the membership function of the conclusion part in each rule i, is a coefficient from membership function of the premise part and the confidence factor, that is related to that rule.

We can provide the preliminary of fuzzy system with complete understanding and knowledge from the quality criteria and the defining input and output linguistic variables with equal terms. After that, we obtain some category of rules by creating fuzzy rules equal to number of terms that are used for defining linguistic variables for every input variable. For the expression of rules, we obtain some categories of rules, by creation one logical mapping between input variable terms in premise part and output variable in conclusion part for every category, that the effect of each rule at ranking should be distinct by the user. This work is done by catching the importance grade of each input quality criterion and located it as a confidence factor related to one category of rules.

For introducing fuzzy rules, we must create a logical mapping according to this point that low or high value of variable is considerable for user. The fuzzy rules for Cost variable that low value is considerable for user can be expressed as follow:

CF_{cost} *IF* S_{Cost}=*very low* *THEN* *Rank=very high*
CF_{cost} *IF* S_{Cost}=*low* *THEN* *Rank=high*
CF_{cost} *IF* S_{Cost}=*moderate* *THEN* *Rank=moderate*
CF_{cost} *IF* S_{Cost}=*high* *THEN* *Rank=low*
CF_{cost} *IF* S_{Cost}=*very high* *THEN* *Rank=very low*

While, we express the fuzzy rules for Cpu-rate variable that high value is considerable for user as follow:

$CF_{Cpu\text{-}r}$ *IF* $S_{Cpu\text{-}r}$=*very high* *THEN* *Rank=very high*
$CF_{Cpu\text{-}r}$ *IF* $S_{Cpu\text{-}r}$=*high* *THEN* *Rank=high*
$CF_{Cpu\text{-}r}$ *IF* $S_{Cpu\text{-}r}$=*moderate* *THEN* *Rank=moderate*
$CF_{Cpu\text{-}r}$ *IF* $S_{Cpu\text{-}r}$=*low* *THEN* *Rank=low*
$CF_{Cpu\text{-}r}$ *IF* $S_{Cpu\text{-}r}$=*very low* *THEN* *Rank=very low*

In fact, having n properties for every resource and defining t term for every variable there are n×t fuzzy rules which are criterion for evaluating different resource properties.

3) Fuzzy-Based Rerource Matchmaking

Matchmaking component needs two input vectors received from the user; the first is property values vector expected to be satisfied by a resource which is selected, and the second one is weight vector which shows the importance of each one of the properties to obtain the final result.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The complete knowledge of resource properties and the definition of linguistic variables and membership functions have a determinative role in fuzzification process of resource properties. The fuzzy rules that expressed based on the logical mapping between terms of linguistic variables, are the criterion for evaluating of different resources. But, these rules are completely neutral. Therefore, the role of user for preferring the rules which express his needs increases. The received user preferences, weight values, are stated as confidence factors of each category of rules.

The fuzzy system uses the proportional of registered properties to the requested properties, according to the equation (18), as input variables. Then the rank of each resource from the optimized resources set is computed. In fact, matchmaking component ranks each of the optimized resources by preferred fuzzy rules and fuzzy inference engine and the specified defuzzification method. According to the computed rank for each resource, a set of resources with the highest rank is selected to send back to the user.

VI. Experimental Results

In order to simulate algorithm we run the GridSim that is a grid java based simulator. Fuzzy system is implemented using FuzzyJ Toolkit [26]. We have also used db4o [27] data base as a repository for advertised resources. We have created ontology of possible resources using protégé API [10], which is a java based API, for semantic description of resources. The structure of the ontology of resources is motivated by the need to provide information about resources. This ontology template has been created according to the basis of Karlsruhe Ontology model [28].

In order to evaluate our algorithm, we have tested it on the 500, 1000, 2000, 4000, 6000, 8000, and 10000 registered resources which are semantically defined according to the ontology template.

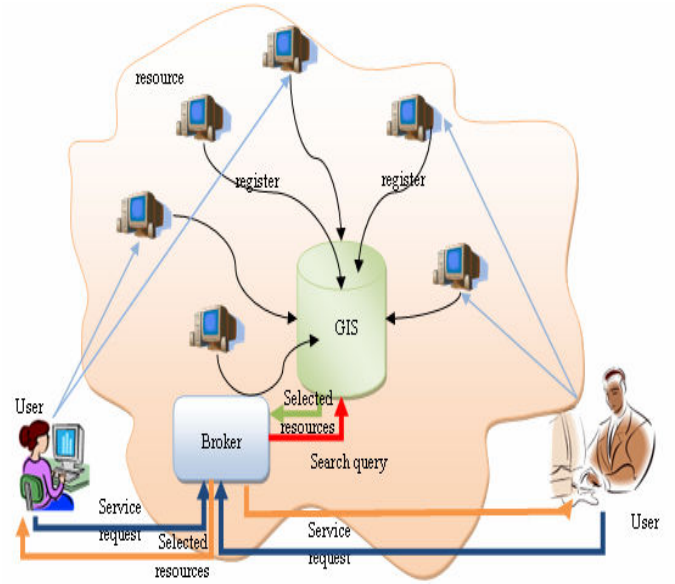


Figure 4. GridSim Architecture

Each resource registers itself in the database as soon as joined the grid by sending its properties which are defined according to the ontology template. For designing Query generator we created users which send resource requests with deferent requested resource properties. Requested resource properties are defined according to the ontology template.

As shown in Fig. 4, user sends its resource query to the GridSim's broker. Broker forwards this request to the Grid Information Server (GIS). The GIS uses the ontology and accesses the database in order to find advertised resources relevant to the requested resource. Retrieved resources ID along with its match degree are sent back to the user.

We have tested our algorithm with resource property certainty of 30%, 50%, 80%, and 100% and for each of these states we have run simulator with deferent number of advertised resources. We have used the average results of the 100 times run of each case for comparison. In first step we show the effect of user preference on the ranking process of resources. Then the effect of our algorithm on the precision and the matchmaking time will be discussed.

For evaluating the precision and the matching time of our algorithm we compared this algorithm with the our previous works; algorithm proposed in [15] which is based on dynamic rough set theory and classical rough set based algorithm proposed in the [16].

A. *User Preferences effect*

In this work, user point of view has direct effect on the computed score for a resource. In fact, user can select a suitable resource through his point of view by assigning weight of every requested property. We have considered the weight of property as the confidence factor which has clear and direct effect on the resource ranking.

Now, with decreasing confidence factor related to one property and fixing others, we can observe that the chart gradient and width of ranking scores in the each case of confidence factor's reduction, is lessen. This subject is true for other properties and is a reason for correct functionality of system according to user preferences. Fig. 5 shows difference of maximum and minimum ranking scores belong to resources against changes of confidence factor related to one property and fixing others in 1.

Several advantages can be stated for this technique. This technique emphasizes on accordance to the user preferences and properties of selected resources. The user clearly states his preferences for selecting the resources which meet its demands. There fore, additional to high care in expression of preferences, for modeling the different user preferences, there is no need to restate the rules according to these different preferences. Also, this technique is extensible against increasing of the properties.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

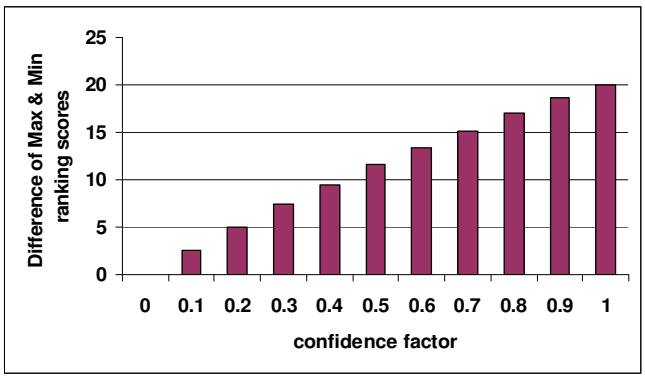


Figure 5. effect of confidence factor changing on width of ranking scores

B. *Precision evaluation*

As mentioned above we test our algorithm with 4 groups of advertised resources. The First group has only 30% properties certainty. The Second group has 50% property certainty and the third group has 80% property certainty and the fourth group has 100% property certainty. Fig. 6 to Fig. 12 shows the comparison of the precision for different numbers of the resources.

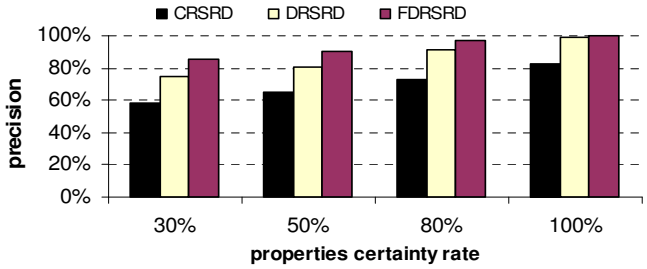


Figure 6. comparison of precision for 500 resources

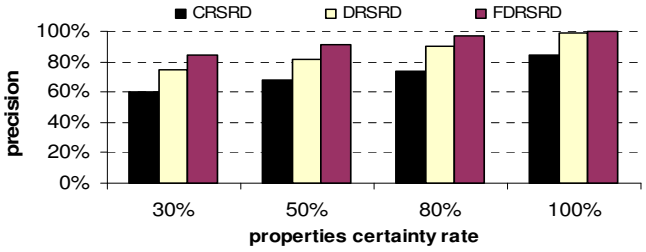


Figure 7. comparison of precision for 1000 resources

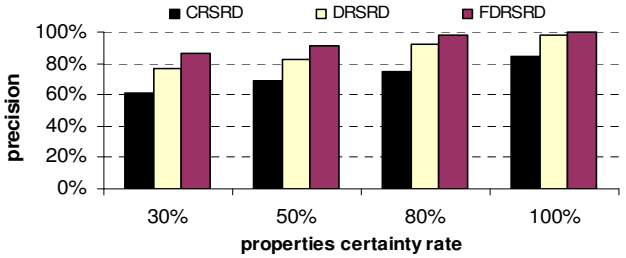


Figure 8. comparison of precision for 2000 resources

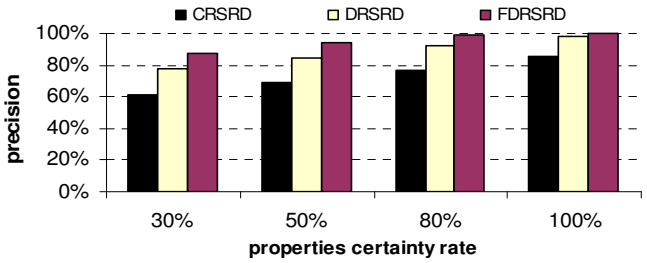


Figure 9. comparison of precision for 4000 resources

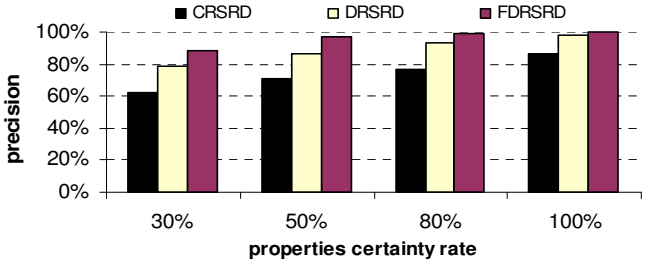


Figure 10. comparison of precision for 6000 resources

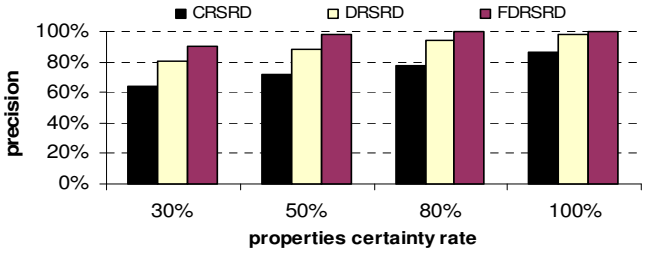


Figure 11. comparison of precision for 8000 resources

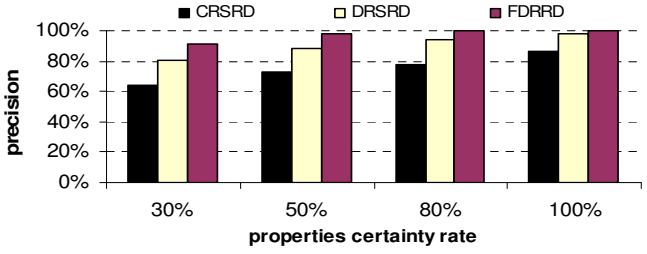


Figure 12. comparison of precision for 10000 resources

As shown in the figures from Fig. 6 to Fig. 12, the precision of the Classic Rough Set Resource Discovery algorithm (CRSRD) is lower than the Dynamic Rough Set Resource Discovery algorithm (DRSRD). This is because of the dynamic properties of the Grid environment. Whereas classic rough set theory can not deal with dynamic properties, classic rough set based algorithm has low precision. But for different rates of certainty, DRSRD is more precise than rough set based algorithm. It is clear that DRSRD has a good effect on dealing with vagueness and dynamic properties of grid. By using fuzzy logic in the matchmaking phase this precision increased more

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

and more. This increment is the result of using fuzzy logic which is a precise method in dealing with complex systems.

Fig. 13 to Fig. 16 show the increment of precision according to the increment of the number of the resources for 30%, 50%, 80%, and 100% certainty rate, respectively. Along with the increase of the number of resources, precision also increases. It is because of the increasing of the survey population.

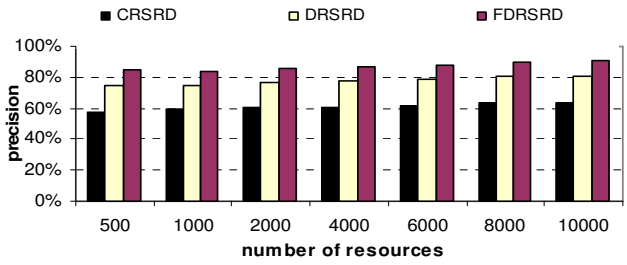


Figure 13. Precision increament for 30% certainty rate

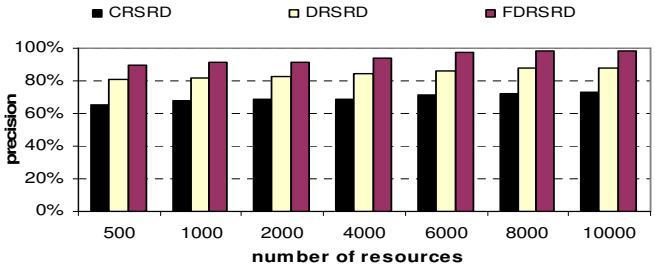


Figure 14. Precision increament for 50% certainty rate

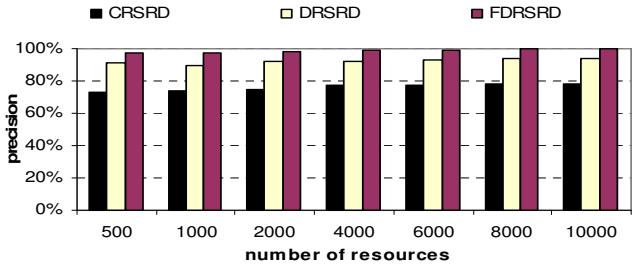


Figure 15. Precision increament for 80% certainty rate

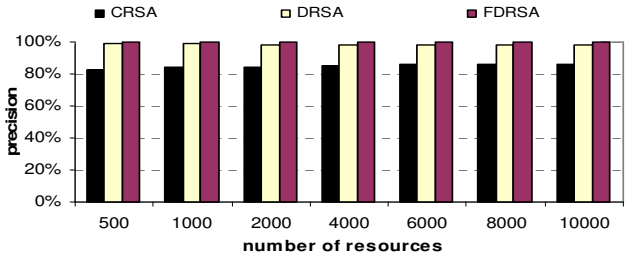


Figure 16. Precision increament for 100% certainty rate

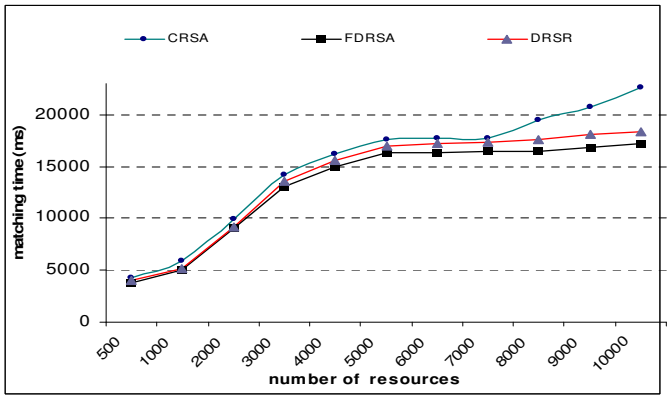


Figure 17. Comparison of the matching time

C. *Matching time evaluation*

For evaluating matching time we run our simulator 100 times with different amount of advertised resources. We have compared the FDRSRD algorithm with the DRSRD and CRSRD algorithms to evaluate the matching time of our algorithm.

Fig. 17 shows that the matching time of FDRSRD algorithm is lower than both the DRSRD and the CRSRD. It is because of the using of the fuzzy logic in the matching phase.

VII. Conclusion And Future Work

Several advantages can be stated for this technique. This technique emphasizes on accordance of the user preferences and the resource properties. The user clearly states his preferences for selecting the resource. There fore, additional to high care in expression of preferences for modeling the different user preferences, there is no need to restate the rules according to these different preferences. Also, this technique is extensible against increasing of the properties.

In this work we have used dynamic rough set theory in order to deal with uncertainty and vagueness existed in the registered resource properties. Experimental results have shown that FDRSRD has good effect in resource discovery process. This effect has shown it self in precision factor which is improved by a acceptable ratio. Furthermore, having a glance on the results revealed that using fuzzy logic along with dynamic rough set lead to the improvement of the precision and the speed up.

Finally, in order to optimize the resource matchmaking phase, it maybe useful to use a technique, such as genetic algorithm in order to rank resources, which in the fitness of the resources is computed according to the approach described in this work.

References

[1] M. Li and M.A.Baker, the Grid Core Technologies, Wiley, 2005.

[2] Open Grid Services Architecture (OGSA), http://www.globus.org/ogsa.

[3] Global Grid Forum (GGF), http://www.ggf.org

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[4] Dong Ya Li, Bao Qing Hu, A Kind of Dynamic Rough Sets, fskd,pp.79-85, Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007) Vol.3, 2007.

[5] Keqiu Li, Deqin Yan, Wenyu Qu, Modifications to Bayesian Rough Set Model and Rough Vague Sets, apscc,pp.544-549, The 2nd IEEE Asia-Pacific Service Computing Conference (APSCC 2007), 2007.

[6] Tian Qiu, Lei Li, Pin Lin, Web Service Discovery with UDDI Based on Semantic Similarity of Service Properties, skg,pp.454-457, Third International Conference on Semantics, Knowledge and Grid (SKG 2007), 2007.

[7] Yue Kou, Ge Yu, Derong Shen, Dong Li, Tiezheng Nie: PS-GIS: personalized and semantics-based grid information services. Infoscale 2007.

[8] BursteinM, Lassila. DAML-S semantic markup for Web services In Proc.of the International Semantic Web Workshop, 2001.

[9] D. Martin, M. Burstein, J. Hobbs, O. Lassila, D. McDermott, S. McIlraith, S. Narayanan, M. Paolucci, B. Parsia, T. Payne, E. Sirin, N. Srinivasan, and K. Sycara, "OWL-S: Semantic Markup for Web Services", http://www.w3.org/Submission/2004/SUBM-OWL-S 20041122/, 2004.

[10] Protégé, http://www. protege.stanford.edu/plugins/owl/.

[11] S. Bechhofer, F. Harmelen, J. Hendler, I. Horrocks, D. McGuinness, P. F. Patel-Schneider, and L. A. Stein. OWL Web Ontology Languag Reference. W3C Recommendation, Feb. 2004.

[12] S. Miles, J. Papay, V. Dialani, M. Luck, K. Decker, T. Payne, and L. Moreau. Personalised Grid Service Discovery. IEE Proceedings Software: Special Issue on Performance Engineering, 150(4):252-256, August 2003.

[13] J. Komorowski, Z. Pawlak, L. Polkowski, and A. Skowron, Rough Sets: a tutorial, Rough Fuzzy Hybridization, Springer, pp. 3-98, 1999.

[14] Li-Xin.Wang, "Course in fuzzy systems and control," 448 pages, Jun 1997.

[15] I. Ataollahi, M. Analoui, "Resource Matchmaking Algorithm using Dynamic Rough Set in Grid Environment",(IJCSIS) International Journal of Computer Science and Information Security, Vol. 4, No.1, 2009, USA.

[16] I. Ataollahi, M. Analoui(in press), "Resource discovery using rough set in Grid environment", 14th International CSI conference (CSICC2009), July 1-2, 2009, Tehran, Iran.

[17] M. Paolucci, T. Kawamura, T. Payne, and K. Sycara. Semantic matching of web service capabilities. Proceedings of 1st International Semantic Web Conference. (ISWC2002), Berlin, 2002.

[18] T. Chen, X. Zhou, N. Xiao, A Semantic-based Grid Service Discovery and Composition, Page(s):527 – 530, Third International Conference on Semantics, Knowledge and Grid, Oct. 2007.

[19] Qi Yong, Qi Saiyu, Zhu Pu, Shen Linfeng, Context-Aware Semantic Web Service Discovery, skg, pp.499-502, Third International Conference on Semantics, Knowledge and Grid (SKG 2007), 2007.

[20] Hau, J., W. Lee, and J. Darlington, a Semantic Similarity Measure for Semantic Web Services, in Web Service Semantics Workshop 2005, Japan.

[21] E. Xu, Shaocheng Tong, Liangshan Shao, Baiqing Ye. Rough Set Approach for Processing Information Table. In Proceeding of SNPD (3)'2007.pp.239~243.

[22] Dong Ya Li , Bao Qing Hu, A Kind of Dynamic Rough Sets, Proceedings of the Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007) Vol.3, p.79-85, August 24-27, 2007.

[23] S.Agarwal and S.Lamparter, "User preference based automated selection of web service composition." in Proceedings of the ICSOC Workshop on Dynamic Web Services, 2005.

[24] I.Sora, D.Todinca, C.Avram, "Translating user preferences into fuzzy rules for the automatic selection of services," 5th International Symposium on Applied Computational Intelligence and Informatics. Timisoara, Romania, May 2009.

[25] D.Nauck and R.Kruse, "How the learning of rule weights affects the interpretability of fuzzy systems," Proc. of 7th IEEE International Conference on Fuzzy Systems, pp. 1235-1240, May 1998.

[26] http://www.iit.nrc.ca/IR_public/fuzzy/fuzzyJTool-kit.html.

[27] http://www.db4o.com/about/productinformation/resources/db4o-4.5-tutorial-java.pdf.

[28] http:\\www.aifb.uni-karlsruhe.de/WBS/sst/Research/Publications/KI-Heft-KAON-Survey-2003.pdf.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Convergence Time Evaluation of AODV and AODV+G in MANETs

Annapurna P Patil Harish.R

Department of Computer Science and Engineering,
M.S. Ramaiah Institute of Technology,Bangalore-54,India.

annapurnap2@yahoo.com wrcharish@gmail.com

Abstract **-Wireless mobile ad-hoc networks are characterized as networks without any physical connections. In these networks there is no fixed topology due to the mobility of nodes, interference, mulitpath propagation and path loss. Hence a dynamic routing protocol is needed for these networks to function properly. Many routing protocols have been developed for accomplishing this task. Selecting most appropriate routing protocol for a particular network scenario is the critical issue.**

Most attempts made at evaluating these algorithms so far have focused on parameters such as throughput, packet delivery ratio, overhead etc. An analysis of the convergence times of these algorithms is still an open issue. The work carried out fills this gap by evaluating the algorithms on the basis of convergence time.

In this paper we present and examine the convergence time evaluation of routing protocols AODV and AODV+G . The algorithm performances are compared by simulating them in ns2. Tcl is used to conduct the simulations, while perl is used to extract data from the simulation output and calculate convergence time. After extensive testing we observed that AODV+G converged well in all situations than AODV. The paper also evaluates the algorithms using the rudimentary metrics-throughput and packet delivery ratio.

Keywords- Routing Protocols, MANETS, Convergence Time.

I. INTRODUCTION

A Mobile Ad-Hoc Network (MANET) is a self-configuring network of mobile nodes connected by wireless links, to form an arbitrary topology. The nodes are free to move randomly. Thus the network's wireless topology may be unpredictable and may change rapidly. Minimal configuration, quick deployment and absence of a central governing authority make ad hoc networks suitable for emergency situations like natural disasters, military conflicts, emergency medical situations etc .

Every device in a MANET is also a router because it is required to forward traffic unrelated to its own use. Almost every year, the world is struck by numerous catastrophic natural disasters, such as earthquake, hurricane, typhoon, tsunami, etc. In such a situation communication systems, fixed or mobile, were usually down due to various reasons. The loss of communication systems as well as information networks made the rescue operation extremely difficult. WiFi-ready notebook PCs(MANET nodes) owned by rescue volunteers themselves to construct a MANET to support such a need.

MANET can be classified based on the communication pattern or the devices used, the variants of MANETs on the type of devices are sensor and ad hoc networks. Routing is one of the critical issue in MANET. Selecting the energy efficient routing protocols improves the performance of the communication. The routing protocols are classified into three types. Proactive protocols maintain routing information for all the destinations, and keep updating this information through periodic updates, an example for this protocol is DSDV[1],OLSR[3]. Reactive protocols don't maintain information for all the destination, rather they discover the route to a destination on demand, an example for this protocol is AODV[2]. Hybrid protocols attempt to combine the advantage of both proactive and reactive protocols, an example for this protocol is TORA[5], ZRP[4], MPOLSR[6]. AODV+G[7] reduces unnecessary traffic will effectively improve the efficiency of those mobile nodes in network.

AODV and AODV+G protocols performs differently under different network scenarios. One protocol might perform better than others in specific situation. These protocols are compared in terms of convergence time to uncover in which situations these types of algorithms have their strengths and weaknesses.

II. RELATED WORK

There are many other works which are related to our work in evaluating routing algorithms. [7] AODV and AODV+G has been compared in terms of Average delay, Packet delivery ratio, Normalized routing load and Routing load reduction, but not in terms of convergence time. [8] AODV and DSDV has been compared in terms of convergence time. Many papers have compared AODV with other routing algorithms. In [9] AODV and DSDV have been compared with average throughput, packet loss ratio, and routing overhead as the evaluation metrics, [10] has compared AODV and DSDV in terms of delay and drop rate, [11] compares AODV and DSDV in terms of throughput, packets received, delay and overload. Similarly, [12] compares

AODV, DSDV and DSR in terms of throughput, delay, drop rate.

III. PROTOCOL SPECIFICATION

This section gives the small presentation of two protocols we evaluate in this paper.

A. AODV

The Ad-hoc On-Demand Distance Vector (AODV) routing protocol is designed for use in ad-hoc mobile networks. AODV is a reactive protocol: the routes are created only when they are needed. It uses traditional routing tables, one entry per destination, and sequence numbers to determine whether routing information is up-to-date and to prevent routing loops.

An important feature of AODV is the maintenance of time-based states in each node: a routingentry not recently used is expired. In case of a route is broken the neighbours can be notified. Route discovery is based on query and reply cycles, and route information is stored in all intermediate nodes along the route in the form of route table entries. The following control packets are used: routing request message (RREQ) is broadcasted by a node requiring a route to another node, routing reply message (RREP) is unicasted back to the source of RREQ, and route error message (RERR) is sent to notify other nodes of the loss of the link. HELLO messages are used for detecting and monitoring links to neighbours.

B. Gossiping & AODV+G

The basic gossiping protocol is simple. A source sends the routing request with probability 1. When a node first receives a routing request, with probability p it broadcasts the request to its neighbors and with probability 1 – p *it* discards the request; if the node receives the same rout request again, it is discarded. Thus, a node broadcasts a given route request at most once. [7] proposes GOSSIP(p,k,m), an extension to the basic gossiping, and suggests that:
A node broadcasts with probability 1 for the first k hops before continuing to gossip with probability p.
If a node with n neighbors receives a message and does not broadcast it, but then does not receive the message from at least m neighbors within a reasonable timeout period, it broadcasts the message to all its neighbors [7].

Hass et al. implements GOSSIP(p,k,m) in Ad Hoc On-Demand Distance Vector protocol (AODV) [18], a typical and well-studied on-demond routing algorithm suited for mobile nodes routing in ad hoc network. We refer this gossip-based AODV as AODV+G. The experiments in [7] shows that gossiping can reduce control traffic up to 35% when compared to flooding and the most significant performance of GOSSIP is achieved by taking p=0.65, k=1 and m=1.

In AODV+G, if the expanding-ring search with a smaller radius fails, rather than flooding to the whole network, here GOSSIP3(.65,1,1) is used. The timeout period of GOSSIP3 should be big enough to allow neighboring nodes to gossip. The NODE_TRAVERSAL_TIME parameter of AODV is a conservative estimate of the average one hop traversal time for packets that includes queuing delays, interrupt processing times and transfer times. GOSSIP3 is not used in the expanding-ring search with a smaller radius, since flooding is more efficient than gossiping for zone with small radius because of the back-propagation effects[17]. The variant of AODV that uses GOSSIP3 is called AODV+G.

IV SIMULATION AND PERFORMANCE ANALYSIS

A. Environment and Assumption

Simulator chosen : The proposed algorithms are simulated on NS2(version 2.33)[13]. NS2 is popularly used in the simulation of routing and multicast protocols, among others, and is heavily used in ad-hoc networking research. ns supports an array of popular network protocols, offering simulation results for wired and wireless networks alike. It can be also used as limited-functionality network emulator. It was necessary to use available implementations of algorithms rather than implement them freshly ourselves, as it is important for the acceptance of an evaluation that the implementation used for evaluation has been scrutinized and accepted as correct by the community. Else the evaluation results will not be accepted as doubt will exist about the correctness of the implementation of the algorithms

Algorithms chosen : Here in this paper we have selected to simulate and evaluate the performance of AODV and AODV+G protocols. AODV is a reactive routing protocol and AODV+G is variant of AODV routing protocol with GOSSIP3. Further experiments can be built based on the results of this project, to compare convergence time performance of algorithms within the same category as well.

Mobility model : The Random Waypoint model is the most commonly used mobility model in research community. At every instant, a node randomly chooses a destination and moves towards it with a velocity chosen randomly from a uniform distribution [0,V_max], where V_max is the maximum allowable velocity for every mobile node. After reaching the destination, the node stops for a duration defined by the 'pause time' parameter. After this duration, it again chooses a random destination and repeats the whole process until the simulation ends.

To create Mobile node Movement Scenario files, the command line that needs to be run under directory : ns-allinone-2.33/ns-2.33/indep-utils/cmu-scen-gen/setdest :
./setdest [-n num_of_nodes] [-p pausetime] [-s maxspeed] [-t simtime] [-x maxx] [-y maxy] > [output-file][14].

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Traffic pattern : Moreover, traffic sources may generate packets at constant bit rate (CBR), or at variable bit rate (VBR). The CBR class is commonly used for voice and data services. In this context, the data rate and the delay remain constant during the packet transmission. More particularly, CBR traffic sources provide a constant flow of data packets of 512 bytes with a transmission rate of 4 packets per second. All CBR traffic scenarios are generated using cbrgen.tcl in NS-2

To create CBR traffic scenario files, under directory :ns-allinone-2.33/ns-2.33/indep-utils/cmu-scen-gen/cbrgen.tcl ./ns cbrgen.tcl [-type cbr|tcp] [-nn nodes] [-seed seed] [-mc connections] [-rate packet/second for one connection] > [output-file][14].

Network scenario : The simulations are conducted using the network simulator ns2 [14]. Random Waypoint mobility model is used. The physical layer simulates the behavior of IEEE 802.11 (as included with ns2). Each node has a radio range of 250 meter, and uses TwoRayGround as the radio propagation model.
All the scenarios are based on the following basic parameters:
cbr (constant bit rate) traffic
topology of size 500 m x 500 m
maximum speed of each node 20 m/s
simulation time 180s
transmission rate (packet rate) 10 m/s
The number of nodes is varied in the range [10,100] in steps of 10 (to represent 10 node densities). Pause time is varied in the range [0,180] in steps of 20 (to represent 10 pause times).

B. Performance Metric

A trace file contains a lot of information which may not be required to analyze the performance of the protocol. We are always interested in some amount of information that is sufficient to predict the efficiency of the protocol. The following performance metric is needed to be taken into consideration in order to analyze and compare the performance of AODV and AODV+G

Convergence Time : In [15], convergence time has been defined as the time between detection of an interface being down, and the time when the new routing information is available. [16] defines a route convergence period as the period that starts when a previously stable route to some destination becomes invalid and ends when the network has obtained a new stable route for. Similarly, we define convergence time as the time between a fault detection, and restoration of new, valid, path information.

[15] calculates convergence time in the IP backbone. The authors arrive at the value of convergence time by deploying entities called 'listeners', which listen to every link state PDU being sent by the is-is protocol. The time when the first 'adjacency down' packet is observed is taken as the time of detection of an interface being down. This failure event is said to end when the listener receives link state PDUs from both ends of the link.

We arrive at the convergence time by measuring the interval between the detection of route failure and successful arrival of a packet at the destination over the newly computed route. This includes not only the routing convergence time, but also the time taken for the packet to traverse the network from the source to the destination over the newly discovered path. Since this is a comparative analysis, and both the routing protocols use shortest distance with number of hops as the metric for distance calculation, both protocols will arrive at the same new route, and the time taken to reach the destination over this new route will be the same (since all physical characteristics are the same). Hence this extra time measured does not affect the comparative analysis.

In any case, the time taken for a packet to travel from the source to the destination is negligible when compared to the time taken for the algorithm to discover the new route, either through route request – route reply sequences as in reactive protocols, or by waiting for an update that contains new route information as in proactive protocols. Also, this automatically verifies that the new path calculated is correct.

The cycle of invalidation of old path and discovery of a new path might occur many times, and for many source-destination pairs over the course of the simulation. Hence the average value of these times is taken as the convergence time of that algorithm for that scenario.
This procedure has been carried out in perl.

Throughput : If y number of packets delivered within t time at a node then the throughput at the node could be defined as y/t. By definition, the throughput needs to be calculated at the bottleneck node, not sender. For the throughput calculation, in general divide the successfully received packets by the simulation time will give the answer. In the trace file there are different levels of received packets such as the RTR or AGT level. The packets received by the node in its AGT level will be the real received packets. Here these packets are filtered from the trace file using perl script.

Packet Delivery Ratio : The ratio between the number of packets successfully received by the application layer of a destination node and the number of packets originated at the application layer of each node for that destination.

V. EXPERIMENTAL RESULTS

Graphs are one of the ways to analyze and compare the results of the trace file. Other methods can also be used for comparison like tabular form showing required output data of the trace file. Simple MS Excel or MATLAB also work for plotting graphs. In this paper the graphs are plotted using xgraph in NS2.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

In order to be able to cover most if not all the types of scenarios the algorithms might face, we varied both the node density (number of nodes) and the node mobility (pause time). The node density (number of nodes) was varied in the range [10,100] in steps of 10 (10 different node densities).

The upper limit of this range was chosen to be 180 because the simulation time is 180s in all the cases. Thus a pause time of 180 implies that the nodes pause in their initial positions for 180 seconds – the entire duration of the simulation. Hence this represents the case where nodes are completely static. Similarly, pause time 0 represents very high mobility where the nodes are in constant motion. Thus we tested each algorithm over 10 node densities x 10 pause times = 100 scenarios.

A. Convergence Time

Convergence time of AODV and AODV+G is calculated using perl script. This script parses trace file created by simulating AODV and AODV+G algorithms to calculate convergence time. In each graph, the node density is fixed and the pause time is varied.

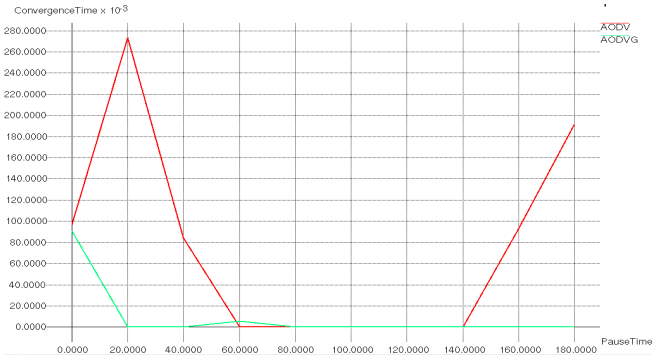


Figure 1: 10 nodes, varying paused time

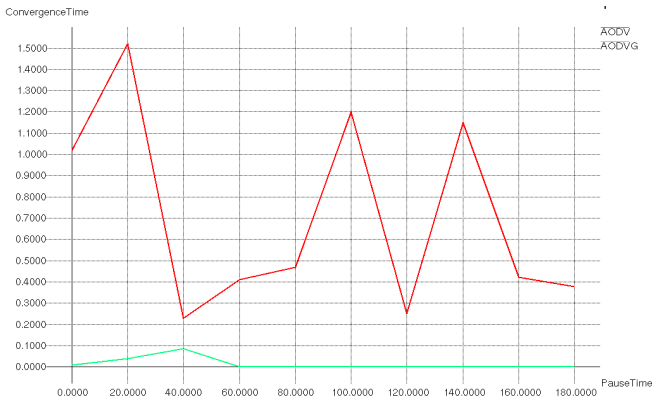


Figure 2: 20 nodes, varying pause time

Figure 3: 30 nodes, varying pause time

Figure 4: 40 nodes, varying pause time

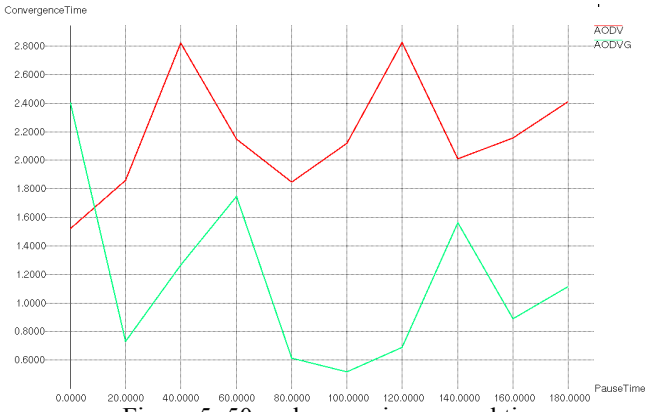


Figure 5: 50 nodes, varying paused time

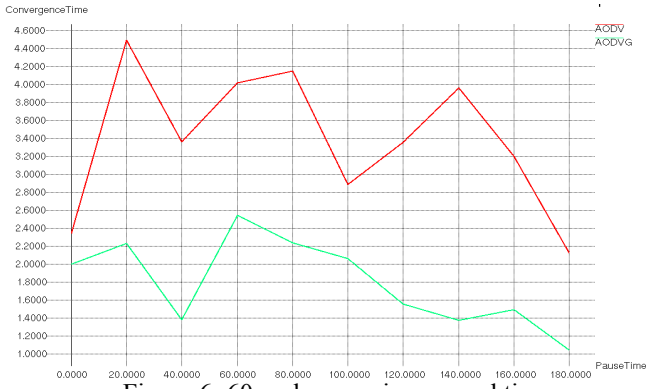


Figure 6: 60 nodes, varying paused time

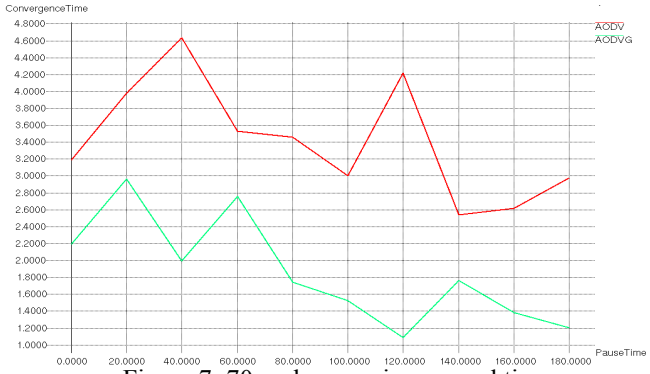


Figure 7: 70 nodes, varying paused time

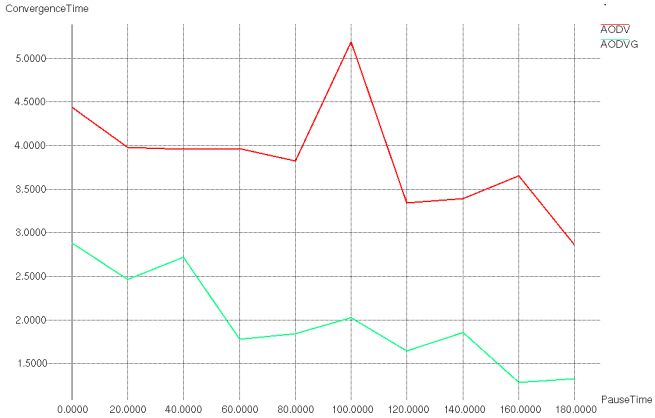


Figure 8: 80 nodes, varying paused time

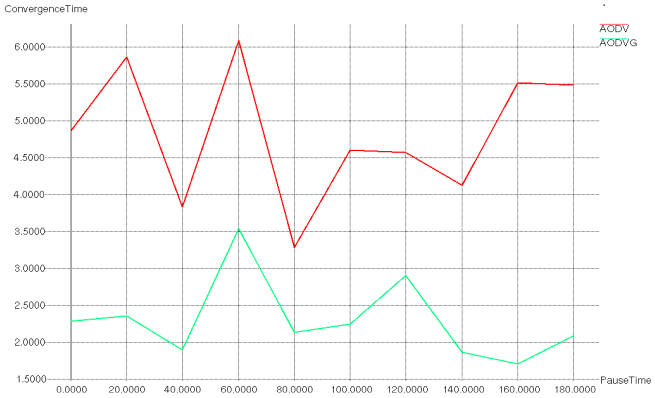


Figure 9: 90 nodes, varying paused time

Figure 10: 100 nodes, varying paused time

Based on the above figures it is found AODV+G convergence time is less than AODV in all assumed network scenarios. It is also found that as the pause time increases the convergence time of AODV+G decreases. Convergence time of both AODV and AODV+G increases as the node density increases.

B.. Throughput

Here node density is varied from 10 to 100 in steps of 10 nodes. For each node density both the algorithms are simulated with varied paused time from 0s ts 180s in steps of 20s. Average throughput in each node density is taken and the graph is plotted as show in figure 12.

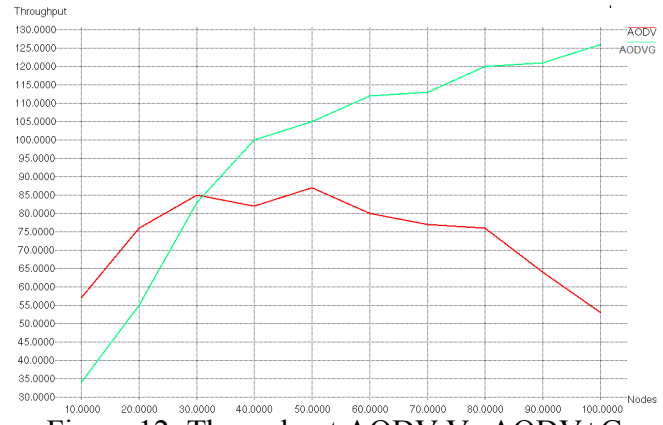


Figure 12: Throughput AODV Vs AODV+G

From the above figure we observed that AODV with the low density performance well than AODV+G. As the node density increases AODV+G throughput increases. With the node density more than 30 nodes the throughput of AODV+G is almost double than AODV.

C. Packet Delivery Ratio

Here node density is varied from 10 to 100 in steps of 10 nodes. For each node density both the algorithms are simulated with varied paused time from 0s ts 180s in steps of 20s. Average of packet delivery ratio in each node density is taken and the graph is plotted as show in figure 11.

Figure 11: Packet Delivery Ratio AODV Vs AODV+G

From the above figure we observed that AODV with the low density performs well than AODV+G. With the node density more than 30 nodes packet delivery ratio of AODV+G is more than AODV.

VI. CONCLUSION

AODV and AODV+G mobile Ad-hoc routing protocols have been presented and evaluated using well know network simulator NS2(version 2.33). AODV+G is gossip based AODV, here GOSSIP3(.65,1,1) is used. These two protocols are evaluated using the network performance metric convergence time. Here we observed that AODV+G converged well than compared to AODV in all assumed network scenarios. We also noticed that with the very low node density throughput and packet delivery ratio of AODV is more than AODV+G. With node density more than 30 nodes AODV+G performs better than AODV. We can extend our work to compare the performance of Adaptive Gossip-based Ad Hoc Routing (AGAR) with Gossip-based Ad Hoc Routing (AODV+G) using convergence time, throughput and packet delivery ratio.

ACKNOWLEDGMENT

We wish to acknowledge our Principal Dr K Rajanikanth, M. S .Ramaiah Institute of Technology,Bangalore-54 and Professor and Head of the Department at CSE Prof .V. Muralidharan for their encouragement which helped us produce this work..

VII. REFERENCES

[1] Charles E.Perkins a nd Pravin Bhagwat, "Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobille Computers " In Proceeding of SIGCOM ' 94 Conference on Communications Architecture, protocols and Applications August 1994.

[2] "Ad hoc On Demand Distance vector (AODV) Routing protocol", RFC 3561, july 2003.

[3] Jing Xie, Luis Girons Quesada and Yuming Jiang "A Threshold-based Hybrid Routing Protocol for MANET" Department of Telematics, Norwegian University of Science and Technology. 2008

[4] Nicklas Beijar, "Zone Routing Protocol (ZRP)", Networking Laboratory, Helsinki University of Technology. 1998

[5] [ns] MANET : TORA Routing Protocol at http://www.mail-archive.com/ns-users@isi.edu/msg05173.html.

[6] Jiazi YI, Sylvain David, Asmaa Adnane, Benoit Parrein, " Multipath-LSR. Simulation and Testpath", 5th OLSR Interop/Workshop, Vienna, Austria (2009)

[7] Z. J. Haas, J. Halpern, and L. Li. Gossip-based ad hoc routing. *IEEE Proceedings of INFOCOM*, 2002.

[8] Annapurna P Patil, Narmada Sambaturu, Krittaya Chunhaviriyakul, "Convergence Time Evaluation of Algorithms in MANETs", (IJCSIS) International Journal of Computer Science and Information security, julyb2009.

[9] R. Khalaf, A. El-Haj-Mahmoud, and A. Kayssi (Lebanon) "Performance Comparison of the AODV and DSDV Routing Protocols in Mobile Ad Hoc Networks", *From Proceeding* (371) Communication Systems and Networks – 2002.

[10] Ebrahim Mahdipour , Ehsan Aminian, Mohammad Torabi, Mehdi Zare, "CBR Performance Evaluation over AODV and DSDV in RW Mobility Model, " *Computer and Automation Engineering, International Conference on*, pp. 238-242, International Conference on Computer and Automation Engineering (iccae 2009), 2009.

[11] Muazzam Ali Khan Khattak, Khalid Iqbal, Prof Dr. Sikandar Hayat Khiyal, " Challenging Ad-Hoc Networks under Reliable & Unreliable Transport with Variable Node Density", Journal of Theoretical and Applied Information Technology, 2005-2008.

[12] REN Wei, YEUNG D.Y, JIN Hai1" TCP performance evaluation over AODV and DSDV in RW and SN mobility models" (School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China) (Department of Computer Science, Hong Kong University of Science and Technology, Hong Kong, China) 2005.

[13] The Network Simulator - ns-2 at http://www.isi.edu/nsnam/ns/

[14] Jaspreet kaur and Cheng Li, " Simulation and analysis of Multicast protocols in Mobile Ad Hoc Networks using NS-2", Memorial University of Newfoundland St.John's, Newfoundland, A1B 3X5, Canada, 2007

[15] G. Iannaccone, C.-N. Chuah, R. Mortier, S. Bhattacharyya, and C. Diot, "Analysis of link failures over an IP backbone," in *ACM SIGCOMM Internet Measurement Workshop (IMW)*, Nov. 2002.

[16] D. Pei, X. Zhao, L. Wang, D. Massey, A. Mankin, S. F. Wu, and L. Zhang, "Improving bgp convergence through consistency assertions, " in *Proc. of IEEE INFOCOM*, 2002.

[17] Zygmunt J. Haas, *Senior Member, IEEE*, Joseph Y. Halpern, *Senior Member, IEEE*, and Li (Erran) Li, *Member, IEEE, "*Gossip-Based Ad Hoc Routing", IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 14, NO. 3, JUNE 2006

AUTHORS PROFILE

The authors are Faculty and Post graduate Student at M S Ramaiah Institute of Technology, Bangalore working in the area of performance evaluation of routing algorithms at the R&D labs, Department of Computer Science and Engineering.

An Improvement of Link Failure Maintenance in Ad-Hoc Wireless Network

Sayid Mohamed Abdule
InterNetWorks Research Group
UUM College of Arts and Sciences
University Utara Malaysia
06010 UUM Sintok, Malaysia
sayidabdule@internetworks.my

Suhaidi Hassan
InterNetWorks Research Group
UUM College of Arts and Sciences
University Utara Malaysia
06010 UUM Sintok, Malaysia
Email: suhaidi@uum.edu.my

Mohammed M. Kadhum
InterNetWorks Research Group
UUM College of Arts and Sciences
University Utara Malaysia
06010 UUM Sintok, Malaysia
Email: kadhum@uum.edu.my

***Abstract*— In mobile ad-hoc wireless network, the link failure is a major challenge. It causes a reduction in performance and efficiency of network resources. This paper presents an enhancement to AODV-DFRP protocol. This developed protocol utilizes the link state prediction method to avoid a link failure in a different way from the existing solutions. The paper focuses only the link failure caused by the nodes mobility, which can be avoided if the routing protocol responses quickly to the network topology change. In this paper, we propose a Local Route Request (LRREQ) mechanism which is expected to show better performance than Local Repair method that is used in standard AODV.** *(Abstract)*

Keyword: Ad-hoc; AODV;

I. INTRODUCTION

Mobile ad-hoc Network (MANET) is one of the most focused research areas in the field of wireless networks as well as mobile communication for the last decade [1] [2]. Ad-hoc network consists of nodes communicating one another with portable radios. In ad-hoc mobile networks, routes are mainly multi-hop because of the limited broadcasting propagation range and frequently, and unpredictably, topology changes, as each network node moves randomly. Therefore, the routing protocol is an essence task as it transferring packets form source node to destination. In other words, it can be described as the process of path finding to reach the desired destination. Finding a new path in ad-hoc network has become a hot research issue. In MANET communication, every node in the network acts as a router as it forwards packet from one node to another.

Many routing protocols have been developed for mobile ad-hoc wireless network. Ad-hoc on-demand distance vector (AODV) is one these protocols which have been used widely.

AODV is a reactive routing protocol of mobile ad-hoc network [3] [4]. When a host or node wants to send a message to destination-node and does not have a valid route to that destination-node, it initiates a route discovery process in order to find out the destination node. Then, it broadcasts a route request (RREQ) packet to its neighbors, which forwards the request to its neighbors until reaches to the desired destination-node or reaches intermediate node which has information about the route to the destination-node [5]. During the route discovery processing, each intermediate node recodes its own sequence number (called broadcast ID). This broadcast ID is incremented for every RREQ that the node initiates, and also records the nodes IP address (source and destination IP addresses). Intermediate nodes can reply the RREQ, in case they have a route to the desired destination-node, only if the destination sequence number is greater than or equal to that contained in the RREQ [6].

In addition, in forwarding process of the RREQ, the intermediate node records the address of the neighbor from which the first copy of RREQ broadcasted is received, to avoid receiving several copies of same RREQ and to avoid looping problem as well. In case of one of the intermediate nodes receives copies of same RREQ, these packets are discarded. Upon receiving the RREQ by the destination-node or intermediate node with recent route to destination, it responds by unicasting a route replay (RREP) packet back to the neighbors that received the first RREQ from [7].

In the presence of Link Failure in AODV, when an intermediate node discovers a link failure in active route, it broadcasts a route error (RERR) packet to inform the source node. Then, the source node will re-initiate a route discovery process again if necessary [8]. One major drawback of AODV protocol is the source route re-initiating upon a link failure; which is the main point that our paper focuses on. In this paper, we come up with a mechanism that avoids the route re-discovery by the source upon link failure and, at the same time, solves the link failure before the current route becomes completely disconnected.

Many ad-hoc routing protocols have been developed, implemented, and they are categorized into different classes. The common routing protocols used in mobile ad-hoc networks (MANETs) are AODV and Dynamic Source Routing (DSR) [9]. These two on-demand ad-hoc network routing protocols are the most widely studied. Subsequently, numerous protocols are being developed or modified which are both proactive and reactive routing protocols, such as Dynamic Destination-Sequenced Distance-Vector routing (DSDV), Optimized Link

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

State Routing (OLSR), Topology Broadcast Based on Reverse Path Forwarding (TBRPF), Signal Stability based Adaptive routing (SSA), and mixed routing protocols such as the Zone Routing Protocol (ZRP), and many others [10]. The goal of this paper is to propose a very useful mechanism to improve AODV-DFRP protocol "in press" [11] and enhance overall ad-hoc wireless network.

II. RELATED WORK

On mobile ad–hoc wireless networks, the route disconnection has a significant negative impact on packet delivery. The common drawbacks and main consequences of route failures are packet dropping, end-to-end delay, and minimum throughput. These result in overall network inefficiencies. In addition, the interval time for link break detection and construction time of alternative path can be high. Therefore, many link predication studies have been proposed, recently, which focus on improving route repair in advance. In [12], a Predictive Preemptive approach has been proposed to route maintenance and to avoid link breakage. The mechanism is initiated when a link is about to be broken instead of waiting for the break to happen. This approach used the Lagrange interpolation mathematical model in order to estimate whether an active route to a neighboring node will fail. When link failure is expected between a node and an upstream neighbor, the upstream neighbor itself attempts, first, to find a route to the destination. If such route is not found within a discovery period, a link failure warning is propagated via upstream nodes to sources that use this link. Source nodes invoke the route discovery mechanism if they need a route to the destination. This approach has two drawbacks: (i) if the problem is happened at the upstream node and not downstream, it is going to take quite a long process to divert route. (ii) The mechanism shows that the unnecessary warning message is still sending to the source after a certain period of time which increases the traffic in the network.

In [13], a new flooding mechanism is suggested to control route paths. The main point of the proposed scheme is to track the destination's location based on the beacon messages of the main route nodes by directional forwarding algorithm. In other words, each node recognizes its location through Global Positioning Systems (GPS) for a one-hop neighbor node, which is at a distance of one hop from the main route, recognizes the location of the main route node through a beacon message. GPS has some issues; such as the cost associated with their use and that they may not work properly because of fading. In addition, other problems with this mechanism could be that the need for synchronization between the internal clocks of nodes [14].

[15] introduces the Packet Received Time (PRT) to predict the link status in order to avoid link breakdown on an active route beforehand. This approach used power measurement of received packets to predict the topological change. This mechanism works as follows; if current signal of the received packet is greater than the previous one, it indicates the link is stable and do not need prediction algorithm. But if it is weaker than the previous received packet's signal power, then prediction algorithm send RERR upstream to locally maintain the route, or to the source (initial node) to establish new RREQ to find a fresh route to the destination. This mechanism has a common weakness as many other mechanisms did, namely sending an unnecessary warning messages propagating back to the source node which again increases the traffic in the network.

III. THE PROPOSED APPROACH

Our approach concerns an avoidance of link breakage before it becomes completely disconnected. The main goal of the proposed mechanism is to avoid sending an unnecessary warning message upon link failure. In addition, it is an improvement of AODV-DFRP protocol proposed by Sayid "in press" [16]. However, the avoidance of the reverse route to the sender (source) is the similar objectives for both LRREQ and AODV-DFRP protocol upon link failure. This paper proposes two mechanisms to improve the DFRP protocol. They are as follows.

- Local Route Request (LRREQ)
- Upstream Notification (UN)

Before we present the new proposed mechanisms, we provide a brief introduction about how the DFRP protocol works upon a link failure. DFRP protocol uses two mechanisms in order to predict the link breakage and take action before it happens. For prediction, it utilizes the link state prediction method to collect the current link signal strength status in order to detect link breakage. For local or neighbors signal status, it utilizes IEEE 802.11 of wireless standards for beacon frames to find the node that has a stronger signal compare to the current connection signal strength among neighbors. Once one of those neighbors replies with positive response and have both stronger signal and route to the destination, the current-node will divert the route into that node.

The weakness of DFRP protocol is, if none of neighbors has stronger signal strength than the current route or it does not have a route to the destination at all, then the protocol will face a problem in order to process the data forwarding.

Our new mechanism, LRREQ, uses the link state prediction method, as well, for predicating an active link. In addition, a LRREQ with beacon mechanism will use one hop neighbors to collect their signal status. LRREQ mechanism uses one hop range. After the link state predication algorithm informs that the link between A and B will be broken soon (Figure 1), node A circulates a local route request to neighbors to check signal status and whether there is a route to the destination as shown in Figure 2.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

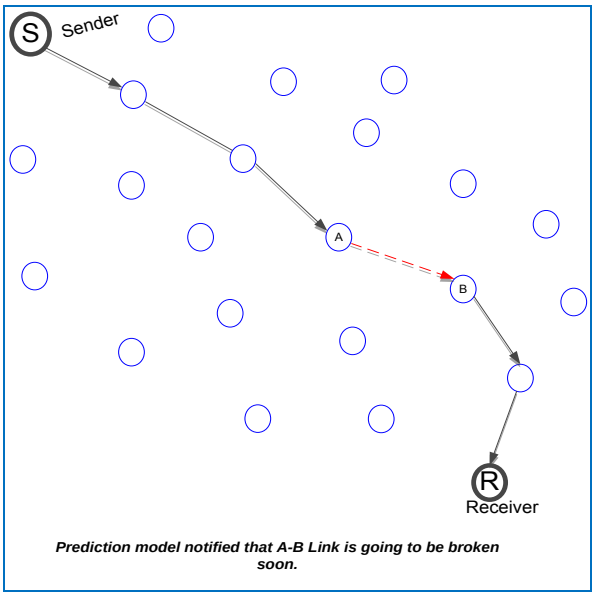


Figure 1. Link A-B is to be broken soon

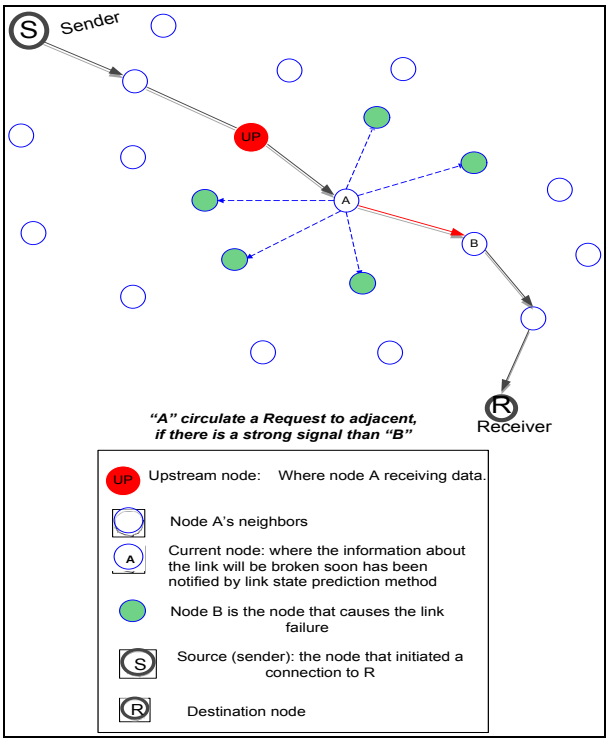


Figure 2. Node A circulates a Request to adjacent

IIII. LOCAL ROUTE REQUEST (LRREQ)

LRREQ comes up with a very useful mechanism on link failure route maintenance. The steps of this mechanism are shown in the through Figures 1-5. LRREQ mechanism is used when the link state prediction model notifies that the link between two intermediate nodes on an active link will be broken soon. This means that, when the A-B link is going be collapsed soon, node A takes an action to sort out the matter, instead of sending warning message back to the sender as in AODV traditional mechanism. In this respect, node A circulates a Local Route Request (LRREQ) among its neighbors, except the node that node A is receiving from (it does NOT circulate to upstream node), to find out if there is a node that has a stronger signal than downstream node (in this case, the downstream node is "B") and, at the same time, has a route to the destination (Figure 2).

A. Upstream Notification (UN)

If current node (in this case, node "A") could not receive a positive response from its neighbors (because no one has a route to the desired destination or maybe they do not have a stronger signal than the current connection), the current node (A) will send a notification to previous node (one hop to upstream) indicating the link will be broken soon.

Thus, when this node (the upstream node) receives such notification from the next hop (downstream), it circulates a Local Route Request into its neighbors except the upstream one and so on. This procedure is presented in Figure 3. The upstream notification continues until a new route to the desired destination is found.

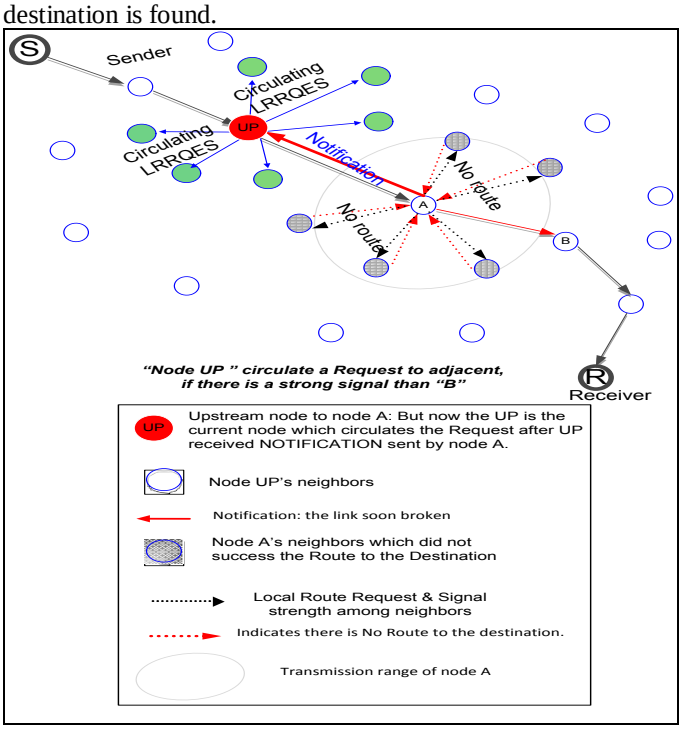


Figure 3. Sending a upstream notification

Figure 3 shows that the node A is failed to sort out the link redirection after the link state prediction informed that the A-B link will not be continued. As mentioned earlier, in LRREQ mechanism, node A will circulates two inquires (for the purpose of finding a stronger signal and a route to destination) to neighbors simultaneously and if one of these inquires not fulfilled, the process will not continue by send a notification to the previous node to indicate that the existing link will be disconnected. This message also includes the status of the signal between nodes A and B. So when the previous node (in this case is node-UP) receives such notification, it will circulate

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

LRREQ to its neighbors, as node-A did. If node-UP not succeed to sort out, it continues sending the notification to upstream nodes until a new route to the destination is discovered.

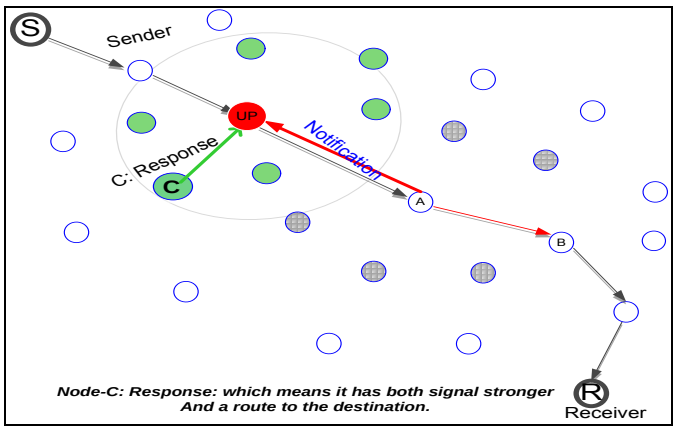


Figure 4. Node C responses to indicate that it has both a stronger signal and route to the destination

If a neighbor node, for example C in Figure 4, has stronger signal than node A to node B, node-UP has to divert the route through the node C (Figure 5), because A-B link is soon to be broken (because node B is moving beyond the coverage range of node A). Intuitively, as the distance between two nodes is greater than the radius of their transmission range, they will not be able to communicate each other.

Figure 5 shows that a new route through node C is constructed. The data flow through node A to node B will continue via the new constructed path without any loss.

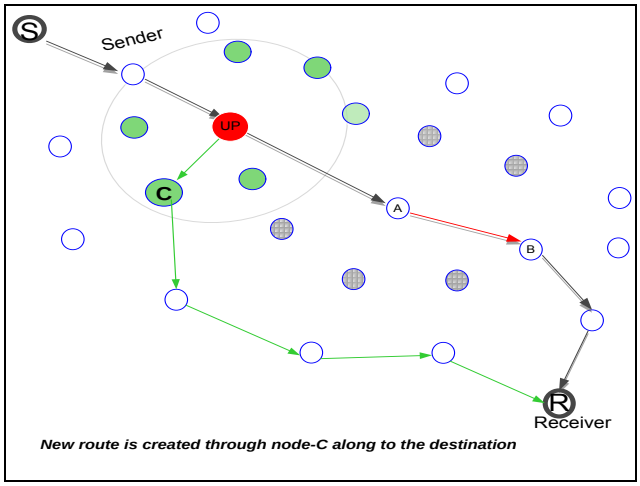


Figure 5. Link Divert Through Node C

IVI. PREDICTION OF AN ACTIVE ROUTE

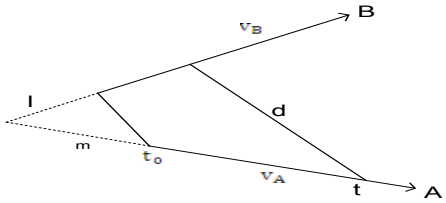
Basically there are two methods, to predict the mobile nodes connectivity.

They are:

- Distance
- Time

In the distance method, the received signal power solely depends on its distance to the transmitter. In this method we assume that the speed, direction, and transmission range are constant. Therefore, the speeds and directions for node A and node B do not change from t_0 to t, where t_0 is the start point. It can calculate the distance d from A to B at time t .As stated in [17] and [18]; the distance can be computed as follows.

$$\mathbf{d^2} = (l + v_A t)^2 + (m + v_B t)^2 - 2\cos\theta(l + v_A t)(m + v_B t) \qquad (1)$$



From this simple calculation, we can predict the time t when mobile node A and B are going to be disconnected. Thus, this method is solely depends on the distance between nodes.

The time method, having knowledge of the distance d, it can predict the time t when node A and node B are going to be disconnected.

We can calculate the distance d from A to B at time t as mentioned earlier. Also we can calculate the remaining time that the link will stay connected with known transmission range r. For example, A and B are within the transmission range r at (x_A, y_A) and (x_B, y_B) moves with speed v_A, v_B at direction θ_A , θ_B. (θ_A and θ_B are in the range of 0 to 2π), respectively. So the remaining time of link connection is:

$$\boldsymbol{T} = \frac{-(ab+cd)+\sqrt{(a^2+c^2)r^2-(ad-bc)^2}}{a^2+c^2}, \qquad (2)$$

Where: $\mathbf{a} = v_A\cos\theta_A - v_B\cos\theta_A$, $\mathbf{b} = X_A - X_B$, $\boldsymbol{c} = v_A\sin\theta_A - v_B\sin\theta_B$, $\mathbf{d} = y_A - y_B$.

VI. ONE-HOP-NEIGHBOR SIGNAL MONITOR

The one hop neighboring signal status is also important to monitor. The circulation of LRREQ among neighbors upon the prediction of a link failure derives creating an array for each node to keep the signal information, called **"signal-status array"**. This array holds three packets for beacon signal strength and their reception time. This is done by *Lagrange Interpolation* method which has the following general definition.

$$L_j(x) := \prod_{\substack{0\le f\le k,\\ f\ne j}} \left(\frac{x-x_f}{x_j-x_f}\right) = \left(\frac{x-x_0}{x_j-x_0}\right)\ldots\left(\frac{x-x_{j-1}}{x_j-x_{j-1}}\right)\left(\frac{x-x_{j+1}}{x_j-x_{j+1}}\right)\ldots \qquad (3)$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$\mathbf{p} = \left(\frac{(t-t_1)\times(t-t_2)}{(t_0-t_1)\times(t_0-t_2)} \times p_0\right) + \left(\frac{(t-t_0)\times(t-t_2)}{(t_1-t_0)\times(t_1-t_2)} \times p_1\right) + \left(\frac{(t-t_0)\times(t-t_1)}{(t_2-t_0)\times(t_2-t_1)} \times p_2\right)$$

The strength powers and time intervals are p_0, p_1, p_2, t_0, t_1 and t_2, respectively. In this respect, if the signal threshold less than the minimum requirement of signal power or there is no route to the desired destination, a warning message is sent to the upstream node. This is shown in the following pseudo code and flowchart.

A. *Pseudo Code of the New Mechanism*

```
// C_acL−sglP  =  Current active Link signal power/*
// T_hold      =  Threshold or required signal power/*
// one-hop-S-Pw  =  one hop neighbors signal power /*
// signal-status-array[3]; = table contains signal status/*

Predict() // execute prediction in every 5 ms
{if (C_acL−sglP < T_hold) than
S-LRREQ(signal-power, dest-Route) /* check one-hop-
neighbor, if one of //them got stronger signal than
current, plus destination //route as well. */
 };
// One-hop-Response
Response()  // one-hop-neighbors response
signal-status-array[3]; // table contains signal //status
{if (one-hop-S-Pw > S-LRREQ(signal-power, dest-
Route)) //* comparing signal power of the one-hop
//neighbor with the active route signal power.*/
Divert-Data (new-path) // Divert data into the new path
Else
Send-UN(signal-power, dest-Route) //* send notification
to upstream node*/
};
```

Figure 6. Pseudo Code

B. *Flowchart of the New Mechanism*

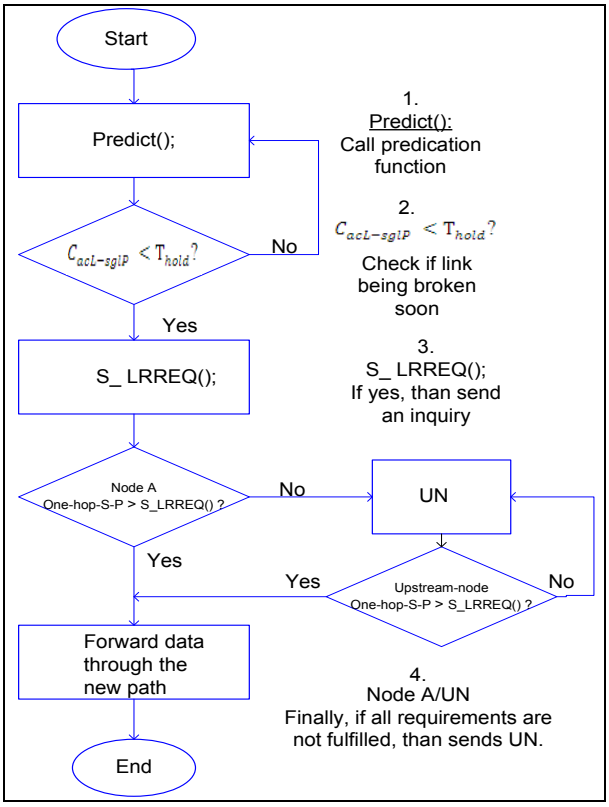


Figure 7. Flowchart

CONCLUSION

In this paper, we propose a new mechanism to improve ad-hoc wireless network performance and to overcome link failure problem. The new mechanism detects link failure before it completely occurs, finds a new route, then divert the data through the new route without loss. This mechanism is supposed to eliminate the route re-establishing by the source upon link failure. In addition, it improves the limitation of DFRP protocol upon receiving a negative response from neighbors (neither signal stronger nor route to the destination). The proposed mechanism is expected to improve the network performance significantly, reduce network overhead, decreases the packet loss, reduce end-to-end packet delays, increase the throughput, utilize the network resources efficiently, and improve packet delivery ratio.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

REFERENCES

[1] Internet Engineering Task Force (IETF) Mobile Ad Hoc Networks (MANET) Working Group Charter, Chaired by Joseph Macker and M. Scott Corson, http: / /www. ietf.org/html. charters/manet-charter. html.

[2] J. Jubin and J.D. Tomow, "The DARPA Packet Radio Network Protocols." Proceedings ofthe IEEE, vol. 75, no. 1, pp. 21-32 January 1987.

[3] C. E. Perkins and E. M. Royer, "The Ad Hoc On-DemandDistance Vector Protocol", Ad hoc Networking, Addison-Wesley, pp. 173-219, 2000.

[4] C. E. Perkins, E. M. Belding-Royer and S. Das, "Ad Hoc On-Demand Distance Vector Routing", IETF RFC3561, July 2003.

[5] [5] Dube R. "Signal stability based adaptive routing for ad hoc mobile networks", IEEE Computer Community, 1997.

[6] Zhijiang C., Georgi G., and Stamatis V. "Routing Protocols for Mobile Adhoc Networks Current Development and Evaluation", Delft University of Technology. Netherlands, 2006.

[7] He, D.; Jiang, S. and Rao, J.,"A Link Availability Prediction Model for Wireless Ad Hoc Networks", Proceedings of the International Workshop on Wireless Networks and Mobile Computing", Taipei, Taiwan, pages D7-D11, April 2000.

[8] Zhijiang C., Georgi G., and Stamatis V. "Routing Protocols for Mobile Adhoc Networks Current Development and Evaluation", Delft University of Technology, Netherlands, 2006.

[9] Awos Khazal Ali. "An Alternative Routing Mechanisms For Mobile Ad-Hoc Networks", Faculty of Computer Science & Information Systems Universiti Teknologi Malaysia, 2008.

[10] Ammar Zahary and Aladdin Ayesh , "Analytical Study to Detect Threshold Number of Efficient Routes in Multipath AODV Extensions" Faculty of Computing Sciences and Engineering De Montfort University Leicester, LE1 9BH, UK, 2007.

[11] Sayid Mohamed Abdule and Suhaidi Hassan, "Divert Failure Route Protocol Based On AODV", proceedings International Conference on Network Applications, Protocols and Services (NetApPS 2010), Universiti Utara Malaysia, 2010.

[12] Hacene et al, "predictive preemptive ad hoc on-demand distance vector routing", Malaysian Journal of Computer Science, Vol. 19(2), 2006.

[13] Byung-Seok Kang et al, "Effective Route Maintenance and Restoration Schemes in Mobile Ad Hoc Networks", Department of Computer Science, Korea Advanced Institute of Science and Technology (KAIST),2010.

[14] Sofiane Boukli Hacene et al, "Predictive Preemptive Ad Hoc On-Demand Distance Vector Routing", Malaysian Journal of Computer Science, Vol. 19(2), 2006.

[15] Naif Alsharabi et al, "Avoid link Breakage in On-Demand Ad-hoc Network Using Packet's Received Time Prediction", College of Computer & Communication Hunan University – changSha , proceedings 19th European Conference on Modeling and Simulation Yuri Merkuryev, Richard Zobel, Eugene Kerckhoffs, 2005.

[16] Sayid Mohamed Abdule and Suhaidi Hassan, "Divert Failure Route Protocol Based On AODV", proceedings International Conference on Network Applications, Protocols and Services (NetApPS 2010), Universiti Utara Malaysia, 2010.

[17] He Dajing, Jiang Shengming and Rao Jianqiang, "A Link Availability Prediction Model for Wireless Ad Hoc Networks", Proceedings of the International Workshop on Wireless Networks and Mobile Computing, pp. D7-D11, Taipei, Taiwan, April 2000.

[18] Rappaport, T. S., .Wireless Communications: Principles and Practice (2nd 107 Edition)., Prentice Hall, ISBN: 0130422320, 2002.

AUTHORS PROFILE

Sayid Mohamed Abdule, received his BSc degree in Computer Science from Agder University, Norway.and, his M.Sc degree in Satellite Communication focusing Quality of Service of VOIP over Satellite from the University Sains Malaysia (School of Computer Science), MALAYSIA. Sayid currently attached to the InterNetWorks Research Group at the UUM College of Arts and Sciences as a doctoral researcher. He is currently pursuing his PhD research in Ad-hoc Mobile networking. His current research interest is on Ad-hoc mobile network routing protocol.

Mohammed M. Kadhum, Ph.D. is an assistant professor in the Graduate Department of Computer Science, Universiti Utara Malaysia (UUM) and is currently attached to the InterNetWorks Research Group at the UUM College of Arts and Sciences as a research advisor. He had completed his PhD research in computer networking at Universiti Utara Malaysia (UUM). His research interest is on Internet Congestion and QoS. He has been awarded with several medals for his outstanding research projects. His professional activity includes being positioned as Technical Program Chair for NetApps2008 and NetApps2010, a technical committee member for various well known journal and international conferences, a speaker for conferences, and a member of several science and technology societies. To date, he has published a number of papers including on well-known and influential international journals.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Associate Professor Dr. Suhaidi Hassan, is currently the Assistant Vice Chancellor of the College of Arts and Sciences, Universiti Utara Malaysia (UUM). He is an associate professor in Computer Systems and Communication Networks and the former Dean of the Faculty of Information Technology, Universiti Utara Malaysia.

Dr. Suhaidi Hassan received his BSc degree in Computer Science from Binghamton University, New York (USA) and his MS degree in Information Science (concentration in Telecommunications and Networks) from the University of Pittsburgh, Pennsylvania (USA). He received his PhD degree in Computing (focussing in Networks Performance Engineering) from the University of Leeds in the United Kingdom.

In 2006, he established the ITU-UUM Asia Pacific Centre of Excellence (ASP CoE) for Rural ICT Development, a human resource development initiative of the Geneva-based International Telecommunication Union (ITU) which serves as the focal point for all rural ICT development initiatives across Asia Pacific region by providing executive training programs, knowledge repositories, R&D and consultancy activities.

Dr. Suhaidi Hassan is a senior member of the Institute of Electrical and Electronic Engineers (IEEE) in which he actively involved in both the IEEE Communications and IEEE Computer societies. He has served as the Vice Chair (2003-2007) of the IEEE Malaysia Computer Society. He also serves as a technical committee for the Malaysian Research and Educational Network (MYREN) and as a Council Member of the Cisco Malaysia Network Academy.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

EFRRA : An Efficient Fault-Resilient-Replica-Algorithm for Content Distribution Networks

Amutharaj. J
Assistant Professor, CSE
Arulmigu Kalasalingam College of Engineering
Krishnankoil, Virudhunagar, INDIA
amutharajj@yahoo.com

Radhakrishnan. S.
Senior Professor, CSE
Arulmigu Kalasalingam College of Engineering
Krishnankoil, Virudhunagar, INDIA
srk@akce.ac.in

***Abstract*— Nowadays, content distribution is an important peer-to-peer application on the Internet that has received considerable research attention. Content distribution applications typically allow personal computers to function in a coordinated manner as a distributed storage medium by contributing, searching, and obtaining digital content. The primary task in CDN is to replicate the contents over several mirrored web servers (i.e., surrogate servers) strategically placed at various locations in order to deal with the flash crowds. Geographically distributing the web servers' facilities is a method commonly used by service providers to improve performance and scalability. Hence, contents in CDN are replicated in many surrogate servers according to some content distribution strategies dictated by the application environment. Devising an efficient and resilient content replication policy is crucial since, the content distribution can be limited by several factors in the network.**

Hence, we propose a novel Efficient Fault Resilient Replica Algorithm (EFRRA) to replicate the content from the origin server to a set of surrogate servers in an efficient and reliable manner. The contributions of this paper are twofold. First we introduce a novel EFRRA distribution policy and theoretically analyze its performance with traditional content replication algorithms. Then, by means of a simulation based performance evaluation, we assess the efficiency and resiliency of the proposed EFRR Algorithm, and compare its performance with traditional content replication algorithms stated in the literature. We demonstrate in experiment that EFRRA significantly reduces the file replication time and maintaining the Delivery ratio as compared with traditional strategies such as sequential unicast, multiple unicast, Fast Replica (FR), Resilient Fast Replica(R-FR), and Tornado codes (TC). This paper also analyzes the performance of sequential unicast, multiple unicast, Fast Replica (FR), Resilient Fast Replica(R-FR), Tornado codes, and EFRRA algorithms in terms of average replication time and maximum replication time.

Keywords- CDN, Fast Replica, Resilient Fast Replica, Efficient Fault Resilient Replica Algorithm, Tornado Codes.

I. INTRODUCTION

Content Delivery Networks (CDNs) [1][2][3] provide services that improve network performance by maximizing bandwidth, accessibility and maintaining correctness through content replication. It offers fast and reliable applications and services by distributing content to cache or edge servers located close to users [1]. A CDN has some combination of content-delivery, request-routing, distribution and accounting infrastructure. The content-delivery infrastructure consists of a set of edge servers (also called surrogates) that deliver copies of content to end-users. The request-routing infrastructure is responsible to directing client request to appropriate edge servers. It also interacts with the distribution infrastructure to keep an up-to-date view of the content stored in the CDN caches.

In particular, Content Delivery Networks (CDNs) optimize content delivery by putting the content closer to the consumer and shorting the delivery path via global networks of strategically placed servers [4]. The CDN's edge servers are the caching servers, and if the requested content is not yet in the cache, this document is pulled from the original server. For large documents, software packages, and media files, push operational mode is preferred. It is desirable to replicate these files at edge servers in advance [6].

While transferring a large file with individual point-to-point connections from an original server can be a viable solution in the case of limited number of mirror server, this method does not scale when the content needs to be replicated across a CDN with thousands of geographically distributed edge replica nodes [5].

This paper is organized as follows. Section II describes the related work. Next, we present the working mechanisms of different content distribution algorithms and our proposed EFRRA content distribution algorithm. Section IV presents discussion on analytical study, experimental results and analyzes the performance of different content distribution algorithms. Finally, the conclusion and future work is summarized.

II. RELATED WORKS

Ludmila et al. [5] proposed a novel algorithm, called Fast Replica, for an efficient and reliable replication of large files in the Internet environment. Instead of downloading the entire file from one server, a user downloads different parts of the same file from different servers in parallel. Once all the parts of the file are received, the user reconstructs the original file by reassembling the different parts. There are several advantages while using a dynamic parallel access. First, as the block size is

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

small, a dynamic parallel access can easily adapt to the changing network/server conditions. Second, as the client is using several connections to different servers, a parallel access is more resilient to congestion and failure in the network/server. Third, the server selection process is eliminated since clients connect to all available servers to retrieve the parts of the file. Fourth, the throughput seen by the client will increase. There is an overhead incurred when opening multiple connections and extra traffic generated to perform block request.

ZhiHui Lu [6] et al proposed Tree Round Robin Replica (TRRR) to improve the work of Fast Replica [5]. They proposed an efficient and reliable replication algorithm for delivering large files in the content delivery networks environment. As part of this study, they carried out some experiments to verify TRRR algorithm in small scale. They demonstrated in experiment that TRRR significantly reduces the file distribution/replication time as compared with traditional policies such as multiple unicast in content delivery networks.

Several content networks attempt to address the performance problem using different mechanisms to improve the Quality of Service (QoS). One approach is to modify the traditional web architecture by improving the web server hardware adding a high-speed processor, more memory, and disk space, or may be with a multi-processor system. This approach is not flexible [7].

In order to offload popular servers and improve end-user experience, copies of popular content are often stored in different locations. With mirror site replication, files from origin server are proactively replicated at surrogate servers with the objective to improve the user perceived Quality of Service (QoS). When a copy of the same file is replicated at multiple surrogate servers, choosing the server that provides the best response time is not trivial and the resulting performance can dramatically vary depending on the server selected [8,9].

Laurent Massoulie [10] proposed an algorithm called the localizer which reduces the network load, helping to evenly balance the number of neighbors of each node in overlay, sharing the load and improving the resilience to random node failures or disconnections. The localizer refines the overlay in a way that reflects geographic locality so as to reduce network load.

In [11], Rodriguez and Biersack studied a dynamic parallel-access scheme to access multiple mirror servers. In their study, a client downloads files from mirror servers residing in a wide area network. They showed that their dynamic parallel downloading scheme achieves significant downloading speedup with respect to a single server scheme. However, they studied only the scenario where one client uses parallel downloading. The authors fail to address the effect and consequences when all clients choose to adopt the same scheme.

J. Kanagharju, J. Roberts, and K.W. Ross[12] studied the problem of optimally replicating objects in CDN servers. In their model, each Internet Autonomous System (AS) is a node with finite storage capacity for replicating objects. The optimization problem is to replicate objects so that when clients fetch objects from the nearest CDN server with the requested object, the average number of ASs traversed is minimized. They showed that this optimization problem is NP complete. They developed four natural heuristics and compared them numerically using real Internet topology data.

Al-Mukaddim Khan Pathan and Rajkumar Buyya [13] presented a comprehensive taxonomy with a broad coverage of CDNs in terms of organizational structure, content distribution mechanisms, request redirection techniques, and performance measurement methodologies. They studied the existing CDNs in terms of their infrastructure, request-routing mechanisms, content replication techniques, load balancing, and cache management. They provided an in depth analysis and state-of-the-art survey of CDNs. Finally, they applied the taxonomy to map various CDNs. The mapping of the taxonomy to the CDNs helps in “gap” analysis in the content networking domain.

James Broberg and Rajkumar Buyya [14] et al proposed MetaCDN a system that exploits ‘Storage Cloud’ resources, creating an integrated overlay network that provides a low cost, high performance CDN for content creators. MetaCDN removes the complexity of dealing with multiple storage providers, by intelligently matching and placing users’ content onto one or many storage providers based on their quality of service, coverage and budget preferences. MetaCDN makes it trivial for content creators and consumers to harness the performance and coverage of numerous ‘Storage Clouds’ by providing a single unified namespace that makes it easy to integrate into origin websites, and is transparent for end-users.

M. O. Rabin [15] and Byers [16] et al proposed an efficient dispersal of information for secure, and fault tolerant data dissemination based on digital fountain approach. The main idea underlying their technique is to take an initial file consisting of *k* packets and generate an *n* packet encoding of the file with the property that the initial file can be restituted from *any k* packet subset of the encoding. For the application of reliable multicast, the source transmits packets from this encoding, and the encoding property ensures that different receivers can recover from different sets of lost packets, provided they receive a sufficiently large subset of the transmission. To enable parallel access to multiple mirror sites, the sources each transmit packets from the same encoding, and the encoding property ensures that a receiver can recover the data once they receive a sufficiently large subset of the transmitted packets, regardless of which server the packets came from. In fact, the benefits and costs of using erasure codes for parallel access to multiple mirror sites are analogous to the benefits and costs of using erasure codes for reliable multicast. For both applications, simple schemes which do not use encoding have substantial drawbacks in terms of complexity, scalability, and in their ability to handle heterogeneity among both senders and receivers.

Ideally, using erasure codes, a receiver could gather an encoded file in parallel from multiple sources, and as soon as *any k* packets arrive from any combination of the sources, the original file could be restituted efficiently. In practice, however, designing a system with this ideal property and with

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

very fast encoding and decoding times appears difficult. Hence, although other erasure codes could be used in this setting, we suggest that a newly developed class of erasure codes called *Tornado codes* are best suited to this application, as they have extremely fast encoding and decoding algorithms [12]. Indeed, previously these codes have been shown to be more effective than standard erasure codes in the setting of reliable multicast transmission of large files [16].

Byers [17] et al proposed a parallel accessing scheme based on tornado codes in which a client is allowed to access a file from multiple mirror sites in parallel to speed up the download. They eliminated complex client-server negotiations and implemented a straightforward approach for developing a feedback-free protocol based on erasure codes. They demonstrated that a protocol using fast Tornado codes can deliver dramatic speedups at the expense of transmitting a moderate number of additional packets into the network. Their scalable solution would be extended to allow multiple clients to access data from multiple mirror sites simultaneously.

Danny Bickson and Dahlia Malkhi [18] proposed a new content distribution network named "Julia" which reduces the overall communication cost, which in turn improves network load balance and reduces the usage of long haul links. Compared with the state-of-the-art BitTorrent content distribution network, the authors found that while Julia achieves slightly slower average finishing times relative to BitTorrent, Julia nevertheless reduces the total communication cost in the network by approximately 33%. Furthermore, the Julia protocol achieves a better load balancing of the network resources, especially over trans-Atlantic links. They evaluated the Julia protocol using real WAN deployment and by extensive simulation. The WAN experimentation was carried over the PlanetLab wide area testbed using over 250 machines. Simulations were performed using the GT-ITM topology generator with 1200 nodes.

Amutharaj J and Radhakrishnan.S [19, 20] constructed an overlay network based on Dominating Set theory to optimize the number of nodes for large data transfer. They investigated the use of Fast Replica algorithm to reduce the content transfer time for replicating the content within the semantic network. A dynamic parallel access scheme is introduced to download a file from different peers in parallel from the Semantic Overlay Network (SON), where the end users can access the members of the SON at the same time, fetching different portions of that file from different peers and reassembling them locally. That is, the load is dynamically shared among all the peers. To eliminate the need for retransmission requests from the end users, an enhanced digital fountain with Tornado codes is applied. Decoding algorithm at the receiver will reconstruct the original content. The authors also found that no feedback mechanisms are needed to ensure reliable delivery. The authors investigated the performance of sequential unicast, multiple unicast and fast replica with tornado content distribution strategies in terms of content replication time and delivery ratio. They also analyzed the impact of dominating set theory for the construction of semantic overlay networks.

Srinivas Shakkottai and, Ramesh Johari,[22] evaluated the benefits of a hybrid system that combines peer-to-peer and a centralized client–server approach against each method acting alone. The key element of their approach is to explicitly model the temporal evolution of demand. They also investigated the relative performance of peer-to-peer and centralized client–server schemes, as well as a hybrid of the two—both from the point of view of consumers as well as the content distributor. They showed how awareness of demand could be used to attain a given average delay target with lowest possible utilization of the central server by using the hybrid scheme.

Zhijia Chen [23] et al addressed the issues related to distributing the media content such as audio, video, and software packages to increasing number of end consumers in higher speed. They integrated Peer-to-Peer (P2P) paradigm into the Internet content distribution infrastructure which provides a disruptive market opportunity to scale the Internet for high quality data delivery. They have done an experimental and analytical performance study over BitTorrent-like P2P networks for accelerating large-scale content distribution over booming Internet. They explored the unique strength of P2P in high speed networks, identified the performance bottlenecks, and investigated and quantified the special requirements in the new scenario, i.e. file piece length and seed capacity. They proposed a Piece-On-Demand (POD) scheme to modify BitTorrent in integration with File system in Userspace (FUSE) with an objective to decrease file distribution time and increase service availability.

Ye Xia [24] et al considered a two-tier content distribution system for distributing massive content and proposed popularity-based file replication techniques within the CDN using multiple hash functions. Their strategy is to set aside a large number of hash functions. When the demand for a file exceeds the overall capacity of the current servers, then the previously unused hash functions were used to obtain a new node ID where the file would be replicated. They developed a set of distributed, robust algorithms to implement the above solutions and evaluated the performance of proposed algorithms.

Yaozhou Ma [25], and Abbas Jamalipour, presented a cooperative cache-based content dissemination framework (CCCDF) to carry out the cooperative soliciting. They investigated two cooperative strategies such as CCCDF (Optimal) with an objective to maximize the overall content delivery performance while CCCDF (Max-Min) with an aim to share the limited network resources among the contents in a Max-Min fairness manner. They demonstrated in simulation results that the enhanced delivery performance was offered by the proposed CCCDF over existing content dissemination schemes.

Oznur Ozkasap [26], MineCaglar and AliAlagoz proposed and designed a peer-to- peer system; SeCond, addressing the distribution of large sized content to a large number of end systems in an efficient manner. It employed a self-organizing epidemic dissemination scheme for state propagation of available blocks and initiation of block transmissions. Their performance study included scalability analysis for different arrival/departure patterns, flash-crowd scenario, overhead analysis, and fairness ratio. Authors studied various performance metrics such as the average file download time,

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

load on the primary seed, uplink/downlink utilization, and communication overhead. They showed that SeCond is a scalable and adaptive protocol which took the heterogeneity of the peers into account.

Jun Wu [27] and Kaliappa Ravindran designed a CDN and studied about the issues related to proxy server placement with an objective to provide its clients with the best available performance while consuming as little resource as possible. They solved surrogate server placement problem as an optimization problem. Among the solutions greedy algorithm yielded better result with low computational cost. The drawback was that it was easy to trap in the local optimum. Jun Wu [27] and Kaliappa Ravindran proposed genetic algorithm to solve this server placement problem. They mathematically modeled the optimization problem and applied genetic algorithm to solve the server placement problem. They conducted the simulation experiments and demonstrated the results with proper justifications for a simple topology for both greedy algorithm and genetic algorithm.

III. CONTENT DISTRIBUTION ALGORITHMS

The CDN consists of many surrogate servers located at different locations which can be clustered or grouped together to form a surrogate server site, so that a client has a good connectivity to at least one of the surrogate servers. A surrogate server site may consist of an array of surrogate servers that cooperate with each other to further enhance the performance of the content delivery network and to avoid any congested paths in the network.

Content from the origin server are proactively replicated at surrogate servers according to some content distribution policies dictated by the application environment. Whole file replication is simple to implement and has a low state cost. It must only maintain the state proportional to the number of replicas. The cost of replicating the entire file in one operation can be cumbersome in both space and time, particularly for systems that support applications with large objects such as audio, video, and software packages. Several content distribution strategies such as Sequential Unicast [5, 19, and 20], Multiple Unicast [5, 19, and 20], and Fast Replica [5, 6, 19, and 20] are already described in the literature.

In this paper, we described the working mechanisms of Optimal Fast Replica (O-FR), and Tornado codes content replication algorithms. Next, we proposed an Efficient Fault Resilient Replica Algorithm (EFRRA) to replicate the content from the origin server to surrogate servers.

A.Working Mechanism of Optimal Fast Replica (O-FR)

The objective of Optimal Fast Replica (O-FR) is to minimize the maximum replication time.

Step 1: Partition the Original file F into 'm' sub files of equal size.

$$\text{Size}(F_i) = \text{Size}(F)/ m \text{ bytes where } 1<=i<=m$$

Step 2: Surrogate server N_0 opens 'm' concurrent connections to surrogate servers $N_1, N_2 \ldots N_m$.

N_0 will send each node N_i the following file and information.

- Surrogate Server list: R = {$N_1, N_2 ,.., N_m$ }(In next step,

  Sub-file F_i will be forwarded to this server list.
- Sub-file F_i.
- Replica amount: k (1<= k <= m).

Step 3 : Every surrogate server N_i { $N_1, N_2, \ldots N_m$ }opens k- concurrent connections and replicate the sub file F_i to the group with k-1 surrogate servers defined in the set { N_j, i<j<i+k, if j<m, then j=(j-1)mod m+1}

In this step, every server N_i = {N1, $N_2 \ldots$ Nm} has the following output links and input links.

- K-1 Output Links : forwarding sub-file F_i to node list { N_j, i<j<i+k, if j<m, then j=(j-1) mod m+1 }
- K-1 Input Links : receiving sub-file F_j from server list { N_j, i-k <j<i, if j<1, then j = j+m }

Step 4 : At last, every node N_i holds k sub files,

$$\{F_j, \text{i-k <j <= i, if j<1, then j = j+m }\}$$

In general case, node list N_i {$N_1, N_2 \ldots$ Nm}, as cache servers and supports concurrent download.

Client Content Request Processing in O-FR: When a user client requests file F from origin server that request will be redirected to the surrogate server list {$N_1, N_2 \ldots N_m$}, and concurrently downloads every sub-file F_i. Then the sub-files will be reassembled in to original file F in the client machine.

In the ideal case, when k=m, every surrogate server N_i holds all of m sub-files of original file F and reorganizes them to form the Original file F in the local node. When the user requests file F from the origin server, the request will be redirected to one surrogate server in the list {N_1, N_2... Nm} and download the whole file F.

B. Working Mechanism of Tornado Codes:

A tornado code [17] is a content distribution strategy based on digital fountain approach [17] which follows block level replication. Block level replication divides each file object into an ordered sequence of finite number of fixed size blocks. This has a benefit of naming the individual parts of an object independently.

*1. Tornado encoding***:** An entire file is fragmented in to a 'k' packets or blocks of equal size and encodes it into 'n' encoded packets where $n=2^A-1$ such that A is the length of the symbol. A random set of blocks of a file will be replicated in multiple surrogate servers. A block-level system may download different parts of an object simultaneously from different nodes and it reduces the overall download time. Since the unit of replication is small and fixed, the cost to replicate an individual block becomes small.

2. Collection step in Tornado: The receiver can run the Tornado decoding algorithm in real-time as the encoding packets arrive, and reconstruct the original file as soon as it determines that sufficiently many packets have arrived.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Assume that any set of n*k encoded packets are received by the receiver, the original file can be reconstructed where 'n' is a small number (i.e.) 1<n<2. The basic principle behind the use of erasure codes is that the original source data, in the form of a sequence of 'k' encoded packets, along with additional redundant packets, are transmitted by the sender and the redundant data can be used to recover lost source data at the receivers. The main benefit of this approach is that different receivers can recover from different lost packets using the same redundant data.

*Advantages of tornado codes***:** No need for retransmission if some encoded packets are lost due to network traffic or physical link outage or due to failure of network components.

Limitations: Some amount of duplicates packets are generated and transmitted through the network when packet loss and parallel download.

The following points need to be considered when determining the size of the blocks requested.

- The number of blocks chosen should be much larger than the number of mirror servers that are accessed in parallel.
- Each block should be small enough to provide a fine granularity of striping and ensure that the transfer of the last block requested from each server terminates at about the same time.
- Each block should be sufficiently large enough to keep the inter block idle time small compared to the download time of a block.
- The document requested via parallel access must be sufficiently large.

C. Working Principle of EFRRA:

A novel algorithm called EFRRA is proposed for an efficient and fault resilient replication of large files in the CDN. Working mechanism of EFRRA can be summarized as follows. In order to replicate a large file among 'n' nodes, the original file is partitioned into 'n' sub files of equal size and each sub file is transferred to a different node in the group. After that, each node propagates its sub file to the remaining nodes in the group. Thus instead of the typical replication of an entire file to 'n' nodes by using 'n' internet paths connecting the original node to the replication group, this replica algorithm exploits n*n diverse internet paths within the replication group where each path is used for transferring $1/n^{th}$ of the file. Hence, the bandwidth requirement is reduced by a factor of 1/n.

Step 1. Distribution of Content to Surrogate Servers

As shown in "Fig. 1," the originator node N_0 opens n concurrent network connections to nodes $\{N_1...N_n\}$, and sends to each recipient node N_i (1 <= i <= n) the following;

- A distribution list of nodes $R = \{N_1...N_n\}$ to which sub file F_i has to be sent on the next step.
- Sub file F_i.

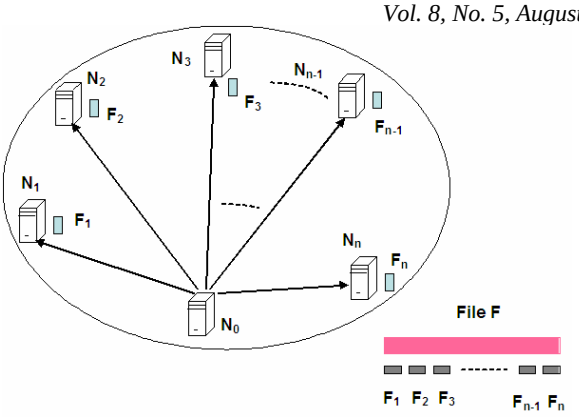


Fig. 1. Distribution Step in EFRRA

Step 2. Adding Fault Resiliency to EFRRA:

It keeps the main structure of the EFRRA replication algorithm practically unchanged while adding the desired property of resilience to node failure. To maintain the resiliency the surrogate servers in the network are exchanging the heartbeat messages with their origin server. The heartbeat messages from surrogate servers to their origin server are augmented with additional information on the corresponding algorithm. Once the content is distributed in the network, the receiver has to recollect all the content from the network in a parallel manner.

For example, if surrogate server N_1 fails during transfer, then this may impact all surrogate servers N_2……N_n in the network because each node depends on node N_1 to receive sub file F_1. In the described scenario as shown in Fig. 2, surrogate server N_i is acting as a recipient server in the replication set. If a surrogate server fails when it acts as the origin surrogate server N_i, this failure impacts all the surrogate servers in the replication group which may be the replication sub tree rooted in surrogate server N_i.

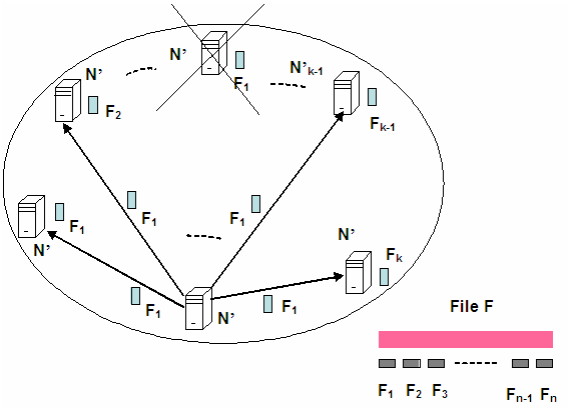


Fig. 2 Adding Resiliency to EFRRA

Step 3. Adding resiliency during Content Collection at the Receiver: Once the entire file is distributed to all the surrogate servers in the overlay network of surrogate servers using step1 then the recipient node or client node has to recollect all the sub files or blocks of the requested file from the overlay network of surrogate servers in a parallel manner. Recipient node retrieves the original source file, in the form of a sequence of 'k'

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

encoded packets, along with additional redundant packets, are transmitted by the sender and the redundant data can be used to recover lost source data at the receivers. Here retransmission of lost packets will not be needed. In this collection step also, EFRRA algorithm maintains resiliency against surrogate server failure and link outages.

IV. PERFORMANCE ANALYSIS

The performance analysis of this paper is twofold. First we theoretically analyze the performance of Efficient Fault Resilient Replica Algorithm (EFRRA) and compare its performance with traditional content replication algorithms. Then, by means of a simulation based performance evaluation, we assess the efficiency and resiliency of the proposed EFRR Algorithm, and compare its performance with traditional content replication algorithms such as sequential unicast, multiple unicast, Fast Replica (FR), Resilient Fast Replica(R-FR), and Tornado codes which were described earlier[19][20].

A. ANALYTICAL STUDY:

Let $Time_i$ denote the transfer time of file F from the origin server N_0 to surrogate server N_i as measured at N_i. Two performance metrics: average and maximum replication times are considered.

Average Replication Time:

$$Time_{avg} = 1/n * \sum_{i=1}^{i=n} \mathbf{Time}_i$$

Maximum Replication Time:

$Time_{Max}$ reflects the time when all the surrogate servers in the overlay network receive k-subfiles (1<=k<=m) of the original file.

$Time_{Max} = \max\{time_i\}$ where i=1..n

In idealistic setting all the nodes and links are homogeneous, and let each node can support 'n' network connections to other nodes at B bytes/sec. Then,

$Time_{distribution} = Size(F) / (nxB)$ (1)

$Time_{collection} = Size (F) / (nxB)$ (2)

a. For Sequential Unicast

Content transfer time is the sum of individual transfer times of entire file from the source to each recipient independently.

$Time_{distribution} = Size (F) / B + Size (F) / B + \ldots + Size (F) / B$

$Time_{distribution} = n * Size (F) / B$

b. For Multiple Unicast

In Multiple Unicast, the original file F is completely replicated concurrently to all nodes $\{N_1,\ldots,N_m\}$ respectively.

$Time_{Distribution} = Size (F) / B$

c. For Fast Replica

$Time_{FR} = Time_{Distribution} + Time_{Collection}$

$Time_{Distribution} = Size (Fi) / B$

$= (1/n)*Size (F) / B$

$= Size (F) /n X B$

$Time_{Collection} = Size (Fi) / B$

$= (1/n)*Size (F) / B$

$= Size (F) /n X B$

Hence $Time_{FR} = 2 x Size (F) / (nxB)$

For Resilient Fast Replica without Node Failure

$Time_{R-FR} = Time_{Distribution} + Time_{RR} + Time_{Collection}$ Where $Time_{RR} = 0$ since Server failure is an occasional event.

$Time_{RR}$ – Time for Resilient Replication

Hence $Time_{R-FR} = Time_{Distribution} + Time_{Collection}$

$= 2 x Size (F) / (nxB)$

For Resilient Fast Replica with Failure of 'm' servers

$Time_{R-FR} = Time_{Distribution} + Time_{RR} + Time_{Collection}$

Hence $Time_{R-FR} = 2 x Size (F) / (nxB) + Time_{RR}$

If m=1, then file F_i has to be replicated to n-1 servers in the overlay. Then $Time_{RR} = Size (F_i) / (n-1) * B$

If m=2, then file F_i has to be replicated to n-2 servers in the overlay. Then

$Time_{RR} = Size (F_i) / (n-1) * B + Size (F_i) / (n-2) * B$

If m='m', then file F_i has to be replicated to n-m servers in the overlay. Then

$Time_{RR} = Size (F_i) / (n-1) * B + Size (F_i) / (n-2) * B + \ldots + Size (F_i) / (n-m) * B$

$= Size (Fi) /B \ (\ 1/(n-1) + 1 / (n-2) + \ldots + (1/(n-m) \ )$

For Large value of 'n' and small value of 'm'

$Time_{RR} = Size(F_i) / B \ (\ 1/n + 1/n + \ldots 1/n)$

$Time_{RR} = Size(F_i) *m / n*B$

$= m* Size(F_i) / n*B$ Where Size (Fi) = 1/n(Size(F))

$= m/n \ (\ size(F)/ \ n*B)$

$Time_{RR} = m/n * Time_{Distribution}$

d. For Optimal Replica

$Time_{O-FR} = Time_{Round1} + Time_{Round2}$

In Round1 Origin server will send File F_i to all the 'n' surrogate servers

$Time_{Round1} = Size (F_i) /n* B$

$= 1/n*Size (F)/ n*B$

$= (1/n*n) Size (F)/B$

In Round2 surrogate servers (N_1, N_2, … N_i) send its sub file F_i to k number of surrogate servers in Round Robin fashion in the Surrogate server set $\{N_1,N_2, .., N_k\}$

Where 1>=k>=n

$Time_{Round2} = Size (F_i) / k*B$ if K>1

$= 1/n*size (F) /k*B$ if k>1

$= 1/n*k (Size (F)/B)$ if k>1

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

$Time_{O\text{-}FR}$ = (1/n*n)*Size (F)/ B + (1/n*k)*(Size (F)/B)

= 1/n (1/n + 1/k) * Size(F) / B

= ((k+n) / n*n*k) * Size (F) / B

Replication Time of Tornado Code algorithm**:**

$Time_{TC} = T_{Encoding} + T_{TR} + T_{Decoding}$ Where $T_{Encoding}$ and $T_{Decoding} \approx T_{TR}$

Where

T_{TR} = (Size ((n*k) encoded packets) / (nXB)) + (Size ((n*k) encoded packets)/ (nXB))

T_{TR} = 2 * (Size ((n*k) encoded packets)/(nXB))

= 2 * (c * Size(F)) / n X B where 1<= c <=2

T_{TR} = 2*c / n * Size(F)) / B where 1<= c <=2

Replication Time of EFRRA Algorithm:

$Time_{Distribution}$ = T_{FR} = (Size (F) + Size (F)) / (nxB) =

= 2 *Size (F) / (nxB)

$Time_{Collection}$ = $T_{TR} + T_{Decoding}$ Where $T_{Decoding} << T_{TR}$

$Time_{Collection}$ = Size ((n*k) encoded packets)

$Time_{EFRRA}$ = (2* Size (F) / (nxB)) + (Size ((n*k) encoded packets) / (nxB))

= (2* Size (F) / (nxB)) + (c * Size(F)) / n X B where 1<= c <=2

$Time_{EFRRA}$ = (2 +c) /n * Size(F)) / B where 1<= c <=2

Therefore, Replication Time proportion of different content distribution algorithms can be expressed as follows

$Time_{SU}$: $Time_{MU}$: $Time_{FR}$: $Time_{R\text{-}FR}$: $Time_{O\text{-}FR}$: $Time_{TC}$: $Time_{EFRRA}$

Size(F) / B **:** 2 x Size (F) / (nxB) **:** m/n (Size(F)/ n*B) : **((k+n) / n*n*k) * Size (F) / B**

1 X Size (F) / B: **2/n**X Size (F) / B: **m/n*n** X (Size (F) / B): **((k+n)/n*n*k)*** Size(F)/B : **2*c / n *** Size(F)) / B **: (2 +c) /n *** Size(F)) / B

n : 1: 2/n : m/n*n : ((k+n) / n*n*k) : 2*c / n : (2 +c) /n where 1<=c<=2

B. EXPERIMENTAL ANALYSIS:

To evaluate the CDN, we used our complete simulation environment, called CDNsim [21], which simulates a main CDN infrastructure. It is based on OMNeT++ library which provides a discrete event simulation environment.

All CDN networking issues, such as surrogate server selection, SON formation, replicating the content from origin server to surrogate servers, implementing the replication algorithms, propagation, queuing, utility computation and pricing computation are computed dynamically via CDNsim [21], which provides a detailed implementation of the TCP/IP protocol, implementing packet switching, packet transmission upon misses etc.

a. Performance of different content distribution schemes in terms of Average Replication Time:

We experimented with 12 different size files; 100 KB, 750 KB, 1.5 MB, 3 MB, 4.5 MB, 6 MB, 7.5 MB, 9 MB, 36 MB, 54 MB, 72 MB, 128 MB and 8 surrogate servers. Fig. 3 shows the average replication time measured by different, individual surrogate servers for different file sizes of 100 KB, 750 KB, 1.5 MB, 3 MB, 4.5 MB, 6 MB, 7.5 MB, 9 MB, 36 MB, 54 MB, 72 MB, 128 MB when 8 surrogate servers are in the replication set. High variability of average replication time under Multiple and Sequential Multicast is identified for larger file sizes. Average Content replication time under *EFRRA* replication algorithm across different file sizes in an 8 surrogate servers replication set is much more stable and predictable. Hence, EFRRA with resilient fast replica for replication and tornado based collection and decoding outperforms most of the cases than sequential multicast, multiple unicast, Fast replica, Resilient Fast Replica(R-FR), Optimal Fast Replica (O-FR), and Tornado codes content distribution schemes.

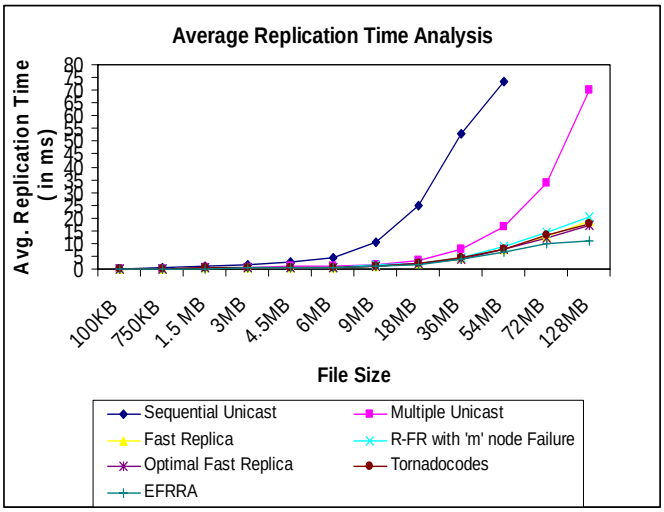


Fig. 3. Average Content Replication Times for various Content distribution algorithms

b. Performance of different content distribution schemes in terms of Maximum Replication Time:

We experimented with 12 different size files; 100 KB, 750 KB, 1.5 MB, 3 MB, 4.5 MB, 6 MB, 7.5 MB, 9 MB, 36 MB, 54 MB, 72 MB, 128 MB and 8 surrogate servers. “Fig. 4,” shows the maximum replication time measured by different, individual recipient nodes for different file sizes of 100 KB, 750 KB, 1.5 MB, 3 MB, 4.5 MB, 6 MB, 7.5 MB, 9 MB, 36 MB, 54 MB, 72 MB,128 MB when 8 surrogate servers are in the replication set. High variability of maximum replication time under Multiple and Sequential Multicast is identified. Maximum File replication time under EFRRA replication algorithm across different file sizes in an 8 surrogate servers replication set are much more stable and predictable.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

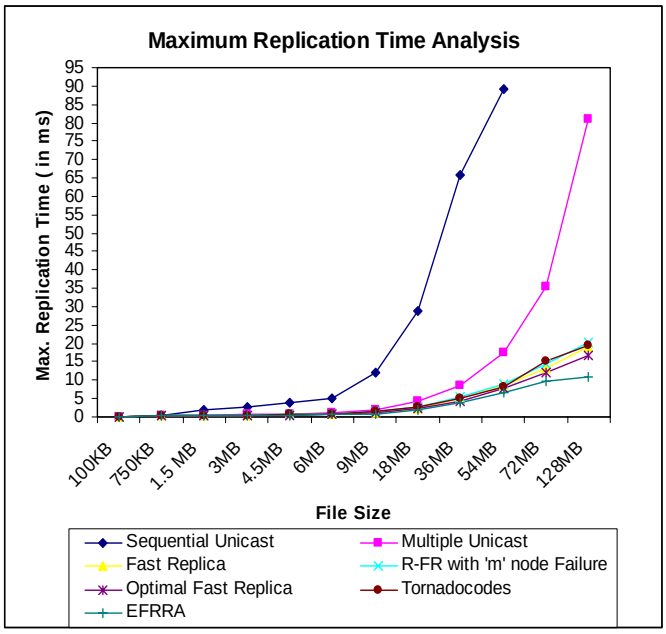


Fig. 4. Maximum Content Replication Times for various Content distribution algorithms

Hence, EFRRA algorithm with resilient fast replica for replication and tornado based collection and decoding algorithm outperforms most of the cases than sequential unicast, multiple unicast, Fast replica, Resilient Fast Replica(R-FR), Optimal Fast Replica (O-FR), and Tornado codes content distribution schemes.

c. Performance of different content distribution schemes during Surrogate server Failure:

The worst case delivery ratio of different content distribution schemes such as sequential unicast, multiple unicast, fast replica(FR), Resilient fast replica(R-FR), Optimal Fast Replica(O-FR), Tornado Codes and EFRRA Content Distribution algorithms when the number of simultaneous surrogate server failures in the CDN has been analyzed and its performance is shown in Fig. 5. The delivery ratio is defined as the ratio of the number of data packets successfully received by the recipient surrogate server to the number of data packets sent by the source surrogate server.

From the delivery ratio analysis shown in Fig.5, we found that the delivery ratio of EFRRA algorithm is consistent during the surrogate server failure. Delivery ratio of traditional algorithms such as Fast Replica (FR), Resilient Fast Replica (R-FR), Optimal Fast Replica (O-FR), and Tornado Codes degrades gracefully with respect to surrogate server failure. It is also observed that delivery ratio of Sequential Unicast and Multiple Unicast content distribution algorithms are degrades drastically with respect to surrogate server failure.

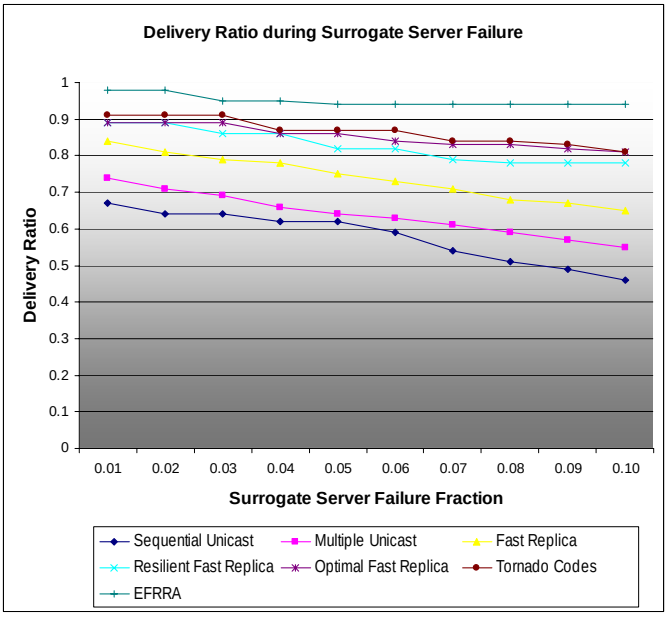


Fig. 5. Delivery Ratio of Content Distribution Algorithms during Surrogate Server Failure.

V. CONCLUSION AND FUTURE SCOPE

In this work, we proposed a novel EFRRA algorithm to distribute the contents in the CDN. We have performed both analytical study and empirical study for analyzing the performance of the proposed EFRRA algorithm and compared its performance with other content distribution algorithms such as Sequential Unicast, Multiple Unicast, Fast Replica (FR), Resilient Fast Replica(R-FR), Optimal Fast Replica (O-FR), and Tornado Codes in terms of average replication time and maximum replication time. We also conducted simulation experiments using CDNsim and analyzed the performance of content distribution algorithms in terms of average content replication time, maximum content replication time and delivery ratio for large files. It is found that EFRRA algorithm outperforms other content distribution algorithms. In future, content distribution algorithms can be compared in terms of other parameters such as surrogate server utilization and optimized network bandwidth usage.

ACKNOWLEDGMENT

The authors would like to thank the Project Coordinator and Project Directors of TIFAC CORE in Network Engineering, Arulmigu Kalasalingam College of Engineering for providing the infrastructure facility in Open Source Technology Laboratory and also thank Kalasalingam Anandam Ammal Charities for providing financial support for this work.

REFERENCES

[1] G. Pallis, and A. Vakali, “Insight and Perspectives for Content Delivery Networks,” Communications of the ACM, Vol. 49, No. 1, ACM Press, NY, USA, pp. 101-106, January 2006.

[2] A. Vakali, and G. Pallis, “Content Delivery Networks: Status and Trends,” IEEE Internet Computing, IEEE Computer Society, pp. 68-74, November-December 2003.

[3] G. Peng, “CDN: Content Distribution Network,” Technical Report TR-125,” Experimental Computer Systems Lab, Department of

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Computer Science, State University of New York, Stony Brook, NY, 2003. http://citeseer.ist.psu.edu/peng03cdn.htm

[4] Content Distribution Networks – State of the art. http://www.telin.nl,2002.

[5] L. Cherksova, J. Kee, "Fast Replica: Efficient Large file Distribution within Content Delivery Networks", Proc. of the 4th SENIX symposium on Internet Technologies, March 2002.

[6] ZhiHui Lu, WeiMing Fu, ShiYong Zhang, YiPing Zhong, "TRRR: A Tree-Round-Robin-Replica Content Replication Algorithm for Improving FastReplica in Content Delivery Networks" in the proceedings of 4th International Conference on Wireless Communications, Networking and Mobile Computing, 2008.

[7] M. Hofmann, and L. R. Beaumont, "Content Networking: Architecture, Protocols, and Practice," Morgan Kaufmann Publishers, San Francisco, CA, USA, pp. 129-134, 2005.

[8] Z. Fei, S. Bhattacharjee, E. W. Zegura and M. H. Ammar, "A novel server selection technique for improving the response time of a replicated service", proceedings in IEEE INFOCOM, vol. 2. San Francisco, CA, March 1998.

[9] M. Sayal, Y. Breibart, P. Scheuermann, and R. Vingralek, "Selection algorithm for replicated web servers," presented at the workshop Internet server performance, Madison, June 1998.

[10] Laurent Massoulie, Anne-Marie Kermarree, Ayalvadi J.Ganesh, "Network Awareness and Failure Resilience in Self – Organising Overlay Networks", Proc. of the 2nd, International Symposium on Reliable Distributed Systems, 2003.

[11] P. Rodriguez and E. Biersack, "Dynamic parallel access to replicated content in the Internet", IEEE/ACM Transactions on Networking, 10(4), Aug. 2002.

[12] J. Kanagharju, J. Roberts, K.W. Ross, "Object Replication Strategies in content distribution networks", Computer Communications, 25(4):367 – 383, March 2002.

[13] A. M. K. Pathan and R. Buyya, "A Taxonomy and Survey of CDNs", Technical Report, GRIDS-TR-2007-4, The University of Melbourne, Australia, Feb. 2007.

[14] J. Broberg, R. Buyya, and Z. Tari, "MetaCDN: Harnessing 'Storage Clouds' for high performance content delivery", Journal of Network and Computer Applications, In Press, Corrected Proof, Available online 5 April 2009.

[15] M. O. Rabin, "Efficient Dispersal of Information for Security, Load Balancing, and Fault Tolerance." In Journal of the ACM, Volume 38, pp. 335- 348, 1989.

[16] J. Byers, M. Luby, M. Mitzenmacher, A. Rege., "A Digital Fountain approach to reliable distribution of bulk data", Proc. of ACM SIGCOMM, 1998.

[17] J. Byers, M. Luby, and M. Mitzenmacher, A. Rege., "Accessing multiple mirror sites in parallel: Using tornado codes to speed up downloads", in Proc. of IEEE INFOCOM, vol. 1, New York, NY, Mar. 1999, pp. 275-283.

[18] Danny Bickson , Dahlia Malkhi, "The julia content distribution network", Proceedings of the 2nd conference on Real, Large Distributed Systems, p.37-41, December 13, 2005, San Francisco, CA.

[19] Amutharaj. J, and Radhakrishnan. S, "Dominating Set based Semantic Overlay Networks for Efficient Content Distribution" proceedings in IEEE ICSCN 2007, vol. 1. Madras Institute of Technology, Anna University, Chennai, Feb 2007.

[20] Amutharaj. J, and Radhakrishnan. S, " Dominating Set Theory based Semantic Overlay Networks for Efficient and Resilient Content Distribution", Journal of Networks, Academy Publishers, Vol 3, March 2008.

[21] K. Stamos, G. Pallis, A. Vakali, D. Katsaros, A. Sidiropoulos, Y. Manolopoulos: "CDNsim: A Simulation Tool for Content Distribution Networks",ACM Transactions on Modeling and Computer Simulation, April 2010.

[22] Srinivas Shakkottai, Ramesh Johari "Demand-Aware Content Distribution on the Internet", IEEE/ACM Transaction on Networking, Vol. 18, No.2. April 2010.

[23] Zhijia Chen, Yang Zhao, Chuang Lin, Qingbo Wang, **"**Accelerating Large-scale Data Distribution in Booming Internet: Effectiveness, Bottlenecks and Practices**",** IEEE Transactions on Consumer Electronics, Vol. 55, No. 2, May 2009.

[24] Ye Xia , Shigang Chen, Chunglae Cho, Vivekanand Korgaonkar, "Algorithms and performance of load-balancing with multiple hash functions in massive content distribution", The International Journal of Computer and Telecommunications Networking**,** Volume 53 , Issue 1, Elsevier**,** January 2009.

[25] Yaozhou Ma, Abbas Jamalipour, "A Cooperative Cache-Based Content Delivery Framework for Intermittently Connected Mobile Ad Hoc Networks", IEEE Transactions on Wireless Communications, Vol. 9, No. 1, January 2010.

[26] Ozkasap O., Caglar M., Alagoz A. "Principles and performance analysis of SeCond: A system for epidemic peer-to-peer content distribution", Journal of Network and Computer Applications Volume 32 , Issue 3 Elsevier, May 2009.

[27] Jun Wu, Kaliappa Ravindran, "Optimization Algorithms for Proxy Server Placement in Content Distribution Networks", IFIP/IEEE International Symposium on Integrated Network Management-Workshops, 2009.

AUTHORS PROFILE

Amutharaj Joyson received his Bachelor of Engineering Degree from Manonmaniam Sundarnar University, Tirunelveli and Master of Engineering from Madurai Kamaraj University, Madurai. He is currently doing his doctoral program from Anna University, Chennai. He is a member of CSI, IAEng, ISTE and Network Technology Group of TIFAC-CORE in Network Engineering. He has published Two papers in International Journals in the Communication Networks, and presented four research papers in International Conferences and Twenty research papers in National Conferences in Network Engineering. His areas of interest are Content Distribution Networks, Mobile Adhoc Networks, Network Technologies, Distributed Computing, Real time Systems, and Evolutionary Optimization.

S. Radhakrishnan received his Master of Technology and Doctorate in Philosophy from Institute of Technology, Banaras Hindu University, Banaras. He is the Director and Head of Computer Science and Engineering Department in Arulmigu Kalasalingam College of Engineering, Srivilliputhur. He is a member of ISTE. He is a Principal Investigator of Research Promotion Scheme (RPS) project funded by Department of Science and Technology, Government of India. He is the Project Director of Network Technology Group of TIFAC-CORE in Network Engineering. This prestigious project is funded by TIFAC, Department of Science and Technology, Government of India and Cisco Systems. He has produced eight Ph. D's and currently guiding twelve Research Scholars in the areas of Network Engineering, Network Processors, Network Security, Sensor Networks, Optical Networks, Wireless Networks, Evolutionary Optimization and Bio-medical Instrumentation.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Comparison of Face Recognition Approach through MPCA Plus LDA and MPCA Plus LPP

Shermina.J
The Department of Computing,
Muscat College, University of Stirling, UK
Muscat, The Sultanate of Oman
Shermina@mctcollege.com

***Abstract* - Face recognition technology has been evolved as an enchanting solution to perform identification and the verification of identity claims. By advancing the feature extraction methods and dimensionality reduction techniques in the application of pattern recognition, a number of face recognition systems has been developed with distinct degrees of success. In this paper, we have presented a biometric face recognition approach based on Multilinear principal component analysis (MPCA) and Locality Preserving Projection (LPP) and the comparison of this to the existing Multilinear principal component analysis (MPCA) and Linear Discriminant Analysis (LDA). This approach consists of four major steps: 1) Face image preprocessing, 2) Dimensionality reduction using MPCA 3) Feature Extraction using LPP and 4) Face recognition using L2 similarity distance measure. Validation of this approach and its comparison with MPCA plus LDA is done with FERET and AT&T face databases. Experimental results show the effectiveness of MPCA plus LPP in comparison with MPCA plus LDA in performance.**

Keywords- Image Processing; Object Recognition; Face Recognition; Image Compression; Multilinear Systems.

I. INTRODUCTION

Biometrics is an automated method of recognizing a person based on their physical or behavioral characteristics. Various types of recognition are available commercially nowadays. Face recognition is the popular area of research in computer vision and the most successful applications of image analysis. It is the biometric identification of a human's face and matching it against a library of known faces. There are two predominant approaches to face recognition, feature based and appearance-based. The geometric feature-based approach uses properties of facial features such as eyes, nose, mouth, chin and their relations for face recognition descriptors. The appearance-based face recognition approach operates directly on image based representation. The whole face region is the raw input to a recognition system and the facial features are processed holistically involving an interdependency between feature and configure of information[10]. The recognition of the face from a still or video clip can be obtained by identifying or verifying one or more persons in the scene using a stored database of faces. The solution includes the process of segmentation of faces (face detection) from cluttered scenes, feature extraction from the face regions, recognition or verification. In identification problems, the system reports the identity of an unknown face from a database of known individuals. In verification problems, the system confirm or reject the claimed identity of the input face[12]. The organization of the paper is as follows: Review of related researches is presented in Section 2. Description about the face recognition methods used in our research is presented in section 3. Methodology of the approach is presented in Section 4. Comparison of the two approaches is given in Section 5. The experimental results is bestowed in section 6 and finally the conclusions are summed up in Section 7.

II. REVIEW OF RELATED WORKS

Many face recognition methods have been developed in the past few decades. Most common feature extraction methods are principal component analysis (PCA) [1], [6] and linear discriminant analysis (LDA) [2], [5], [7], [8]. Another linear technique which is used for face recognition is Locality Preserving Projections (LPP) [3], [4], which finds an embedding that preserves local information, and gains a face subspace that best detects the essential face manifold structure. PCA is a linear method for unsupervised dimensionality reduction. [26]. Kirby and Sirovich used PCA in 1987 in order to obtain a reduced image[27]. Turk, Pentland, Moghaddam and Starner extended the idea to eigenspace projections in the solution of face recognition[6]. PCA on tensor objects requires their reshaping into vectors in a very high-dimensional space, which not only results in high computational and memory demands but also breaks the natural structure and correlation in the original data. Useful representations can be obtained from the original form and PCA extensions operating directly on the tensor objects rather than their vectorized versions are emerging

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

recently [11]. Mittal, N. and Walia, E. [17] introduced a fast PCA based face recognition algorithm. The efficiency of the proposed algorithm is verified using Indian face database, and the gained results prove an improvement in performance of the proposed algorithm as compared to the same with PCA method. Hossein Sahoolizadeh *et al.* [18] have presented a face recognition method on the basis of principal Component Analysis, Linear Discriminant Analysis and neural networks. The proposed method is investigated on Yale face database. Experimental results on this database demonstrated the effectiveness of the proposed method for face recognition with a decrease number of misclassification while comparing with earlier methods. A two-dimensional PCA (2DPCA) is proposed in [28] that constructs an image covariance matrix using image matrices as inputs. However, linear transformation is applied only to the right side of image matrices so the image data is projected in one mode only, resulting in poor dimensionality reduction. A more general algorithm named generalized low rank approximation of matrices (GLRAM) was introduced in [31], which applies two linear transforms to both the left and right sides of input image matrices and results in a better dimensionality reduction than 2DPCA. GLRAM is developed from the perspective of approximation while the generalized PCA (GPCA) is proposed in [29] from the view of variation maximization, as an extension of PCA. Later, the concurrent subspaces analysis (CSA) is formulated in [30] for optimal reconstruction of general tensor objects, which can be considered as a generalization of GLRAM, and the multilinear PCA (MPCA) introduced in [11] targets at variation maximization for general tensor objects in the extension of PCA to the multilinear case, which can be considered as a further generalization of GPCA[26]. Wangmeng Zuo *et al.* [32] have depicted two LDA-based methods, post-processed Fisherfaces and bi-directional PCA plus LDA (BDPCA+LDA). pFisherfaces utilizes 2D-Gaussian filter to smooth classical Fisherfaces, and BDPCA+LDA is a LDA accomplished in the BDPCA subspace. Then they introduced a combination framework where they merged these two LDA-based approaches. Two popular face databases, the ORL and the FERET, are being used to compute the efficiency of the combination framework. The outcome of the experiments revealed that the combination framework is superordinate to pFisherfaces, BDPCA+LDA, and other appearance-based methods in terms of recognition accuracy. Deng Cai *et al.* [19] have proposed an appearance based face recognition method, called orthogonal Laplacian face in which face data may be generated by sampling a probability distribution that has support on or near a sub-manifold of ambient space. This algorithm is based on locality preserving projection(LPP) algorithm. Experimental results on three face databases clearly demonstrated the efficiency of the proposed algorithm. A unique feature extraction technique is not felicitous when the dimensionality of face images attempts to reach its peek and therefore an analytical study made using the combination of two feature extraction methods on distinctly separate subspaces seems to be more effective in the performance of face recognition.

III. REOGNITION METHODS

Image based face recognition techniques can generally be divided into two groups: 1) appearance-based that uses a holistic texture features which can be applied to either whole-face or particular regions in a face image, and 2) feature based, that uses geometric facial features like mouth, eyes, brows, cheeks etc. and the geometric relationships existing between them [15]. Few appearance-based methods are Principal Component Analysis (PCA), Independent Component Analysis (ICA), Linear Discriminant Analysis (LDA) and Locality Preserving Projections (LPP). Among these we have utilized the appearance-based recognition methods such as MPCA, LDA and LPP.

A. MULTILINEAR PRINCIPAL COMPONENT ANALYSIS (MPCA)

Principal Component Analysis (PCA) is the widely used technique in face recognition which is to reduce the dimensionality of a image set consisting of a large number of interrelated variables, by retaining most of the variations in the image set. This is accomplished by transforming to a new set of variables, called the principal components which are uncorrelated but ordered so that the first few retain most of the variation present in all of the original variables. This method of linear algebra addresses single-factor variations in facial images. If factors like lighting, viewpoint, expression, are also permitted to modify facial images, eigenfaces undergoes difficulty. Multilinear Principal Component Analysis (MPCA) is the extension of PCA that employs multilinear algebra and proficient of learning the interactions of the multiple factors[11].

B. LINEAR DISCRIMINANT ANALYSIS (LDA)

Linear Discriminant Analysis (LDA) is a classical supervised linear subspace learning method for feature extraction and dimension reduction that has been very successful and applied in various applications. It aims to derive the most discriminative features and produces a class specific feature space based on the maximization of Fisher's Discriminant Criterion, which is defined as the ratio of betweenclass scatter to withinclass scatter. Classical LDA projects the data onto a lower-dimensional vector space such that the ratio of the between-class distance to the within-class distance is maximized, thus achieving maximum discrimination. The optimal transformation can be obtained by applying the eigen decomposition on the scatter matrices[7].

C. LOCALITY PRESERVING PROJECTIONS (LPP)

Locality Preserving Projection (LPP) is one of the linear approximations of the nonlinear Laplacian Eigenmap. The locality preserving quality of LPP is responsible to be of particular use in information retrieval applications. In order to retrieve audio, video, text documents under a

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

vector space model, then it ultimately needs to do a nearest neighbor search in the low dimensional space. Since LPP is intended to preserve only local structure, it is probable that a nearest neighbor search in the low dimensional space will produce similar results compared to that in the high dimensional space. This allows an indexing scheme to access quickly [3].

IV. METHODOLOGY

In face recognition problems, gallery and the probe are the two types of the data sets used. The gallery set contains the set of data samples of unknown identities and is used for training. The probe set is the testing set where the data samples of unknown identity are to be identified and classified by matching with corresponding entries in the gallery set. MPCA accepts tensor samples of the dimension for feature extraction[11]. However, in practice tensor object samples are of different dimensions. Therefore, the input tensors need to be normalized to standard dimensions and then preprocessing procedure is used to remove the noise, if required. The normalized tensor samples are then centered by subtracting the mean obtained from the gallery tensors. From the gallery set, a set of eigentensors is obtained with reduced dimensionality and each entry of a projected tensor feature can be viewed as a scalar feature corresponding to a particular eigentensor. Some of the small variation and noise are removed in the projection. MPCA is an unsupervised technique without taking class labels into account. Hence, variation captured in the projected tensor subspace includes both the within-class variation and the between-class variation. A larger between-class variation to the within-class variation indicates good class separability. Hence the feature selection strategy selects eigentensors according to class discrimination power which is the ratio of between-class variation over within-class variation. In classification[2], the distance between feature vectors is of importance as it determines the performance of the classification module. Distance between unknown probe image and gallery images stored in database are measured by any of the distance measures L1, L2, angle or Mahalanobis distance.

A. MPCA PLUS LPP APPROACH

Image processing techniques such as normalization and resizing of the face images are employed in preprocessing in order to improve the face image. Next, the feature extraction is achieved by merging the MPCA along with LPP to calculate the feature projection matrices. The feature projection matrices project the face images to lower dimension and generate the face feature vectors. While testing, the test face sample is mapped onto a feature matrix with the assistance of MPCA and LPP. The face recognition can be done by comparing the test feature matrix with the enrolled face features in the database using L2 distance. When processing a face, the features like variations in light, image quality, persons' pose, facial expressions and more should be taken into account. To identify correct individuals successfully these variations must be related by some way to provide a valid answer. Feature extraction or dimensionality reduction is a methodology to transform a high-dimensional data set into a low-dimensional equivalent representation that assumes to retain most of the information regarding the underlying structure or the original physical phenomenon [16]. The main tendency of using feature extraction is its representation of data in a lower dimensional space that computes through a linear or non-linear transformation satisfying certain properties. MPCA is a multilinear subspace learning method that extracts features directly from multi-dimensional objects. MPCA receives the set of face image samples of the same dimensions as input for feature extraction. The resultant output of the MPCA is the dimensionally reduced feature projection matrix of face images. Locality Preserving Projection (LPP) is one of the linear approximation obtained from the nonlinear Laplacian Eigenmap [9]. The dimension reduced feature projection matrices of face image samples obtained using MPCA is then fed as an input to the LPP algorithm. The dimensional reduced feature matrices of the training sample images obtained using the MPCA and LPP techniques are stored in a database. While we are testing the face images, the aforesaid techniques are applied to generate the feature matrix and thereby a similarity measure is carried out on the sample face images. Various face recognition systems may use different distance measures while matching query images with the nearest database images. Our Face recognition approach used here is performed using L2 distance measure. The L2 distance is computed between the face images present in the database and the query image for matching process. The similarity distance measure for a pair of face images is computed in which a threshold determines whether the face pair is classified as same or different.

B. MPCA PLUS LDA APPROACH

The use of Linear Discriminant Analysis to perform feature extraction combined with the projected multilinear arrays produces the approach MPCA+LDA. The feature vector can be used directly for recognition, and a classical linear discriminant analysis (LDA) can also be applied to obtain an MPCA+LDA approach for recognition [23], similar to the PCA+LDA approach. MPCA+LDA extend PCA to use the entire face image as a tensor object inorder to preserve the relationship between neighboring pixels. LDA make use of class information in data and attempts to minimize the within class scatter and to maximize the between class scatter. When using the AR database, the recognition rate is proved to be better for full image face.

V. COMPARISON OF MPCA PLUS LDA WITH MPCA PLUS LPP

The proposed approaches of MPCA+LDA and MPCA+LPP is tested using the FERET database [13] and AT&T database [21] of faces. Performance is measured by procedures applied to FERET facial images and AT&T facial images. In particular, all the images were

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

preprocessed using normalization and then by resizing of the images. The images are divided into training set and test set. The training set is used to initialize and prepare the system to recognize the arbitrary images. The test set is the set of images used to evaluate the performance of the system once the training phase is completed. Here, in FERET database, we use nearly 80 images for training and 160 images for testing. In AT&T database, we use 100 images for training and 200 images for testing process. The experiments were conducted using the L2 distance measure. The common way to measure the biometric recognition accuracy is to compute the false acceptance rate (FAR) and false rejection rate (FRR). FAR is the percentage of incorrect acceptances. FRR is the percentage of incorrect rejections. The accuracy measurement of the overall approach is computed as *100-(FAR/FRR)/2*. Genuine acceptance rate (GAR) is an overall accuracy measurement of the approach.

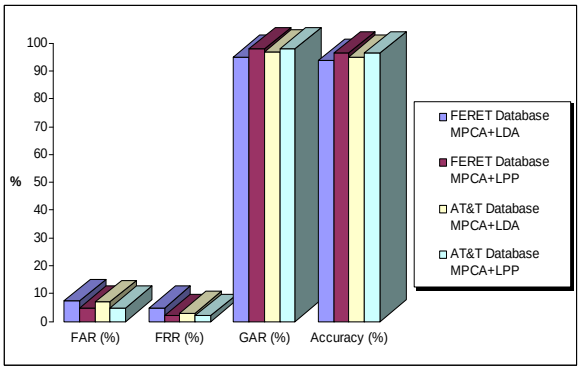
VI. EXPERIMENTAL RESULTS

The percentage of accuracy is given in the following table 1.

Table 1: Accuracy rates on FERET and AT&T databases

Database	**Face Recognition Approach**	
	MPCA+LDA	*MPCA+LPP*
FERET	93.80%	96.50%
AT&T	95%	96.50%

The comparative results are analyzed with the recognition rates of the feature extraction methods such as MPCA plus LDA and MPCA plus LPP by means of charts. The charts clearly shows that the recognition performance of the approach MPCA plus LPP is more efficient when compared to MPCA plus LDA. The face recognition results of MPCA+LPP is compared with MPCA+LDA based on the measures such as false acceptance rate, false rejection rate, genuine acceptance rate and recognition accuracy. The comparative result of each recognition rate is shown in the following chart.



VII. CONCLUSION

Face Recognition is the active area of research over the past two decades which have resulted in successful techniques for recognizing the facial images. In this article, the comparison of Face recognition approaches MPCA plus LDA and MPCA plus LPP is presented. The combined appearance based technique such as MPCA and LPP yield to produce a high face recognition rate compared to MPCA and LDA technique. Experimental results on FERET and AT&T database demonstrated the effectiveness of the proposed approach with improved recognition accuracy. In future, fusion of MPCA plus other face recognition approaches could be experimented and the comparitive analysis could be done to verify the performance of the approach.

ACKNOWLEDGMENT

I would like to thank Dr.V.Vasudevan for providing the guidance in the preliminary discussions on the comparison of the two approaches used in order to proceed further in my area of research.

REFERENCES

[1] H. Moon, P.J. Phillips, “Computational and Performance aspects of PCA based Face Recognition Algorithms”, Perception, vol. 30, pp. 303-321, 2001.
[2] P.N. Belhumeur, J.P. Hepanha, and D.J. Kriegman, “Eigen faces vs.fisherfaces: recognition using class specific linear projection,”IEEE. Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, 1997.
[3] X. He and P. Niyogi, “Locality preserving projections, ”in Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2003.
[4] Li, Ruidong; Zhu, Lei; Zou, Shaofang; “Face Recognition Based on an Alternative Formulation of Orthogonal LPP”, Control and Automation, 2007, IEEE International Conference on May 30 2007-June 1, pp.2374 – 2376, 2007.
[5] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, No. 7, pp. 711-720, 1997.
[6] M. Turk and Pentland, “Face Recognition Using Eigenfaces”, in Proc. IEEE International Conference on Computer Vision and Pattern Recognition, Maui, Hawaii, 1991.
[7] J. Yang, Y. Yu, and W. Kunz, “An Efficient LDA Algorithm for Face Recognition”, The sixth International Conference on Control, Automation, Robotics and Vision, Singapore, 2000.
[8] W. Zhao, R. Chellappa, and P. J. Phillips, “Subspace Linear Discriminant Analysis for Face Recognition”, Technical Report CAR-TR-914, Center for Automation Research, University of Maryland, 1999.
[9] M. Belkin and P. Niyogi, “Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering,” Advances in Neural Information Processing Systems 14, Vancouver, British Columbia, Canada, 2002.
[10] Borut Batagelj and Franc Solina, “ Face Recognition in Different Subspaces: A Comparative Study”
[11] Haiping Lu, K.N. Plataniotis, A.N. Venetsanopoulos, "MPCA: Multilinear Principal Component Analysis of Tensor Objects", IEEE Transactions on Neural Networks, Vol. 19, No. 1, January 2008.
[12] Zhao, W. Chellappa, R., Phillips, P. J. and Rosenfeld, A., “Face Recognition: A Literature Survey”, ACM Computing Survey, December, pp: 399-458, 2003.
[13] P.J. Phillips, H. Moon, P. Rizvi, and P. Rauss, ˝The FERET evaluation method for face recognition algorithms, ˝ IEEE Transactions on Pattern Analysis and Machine Learning, vol. 22, pp. 1090-1104, 2000.
[14] S. T. Gandhe, K. T. Talele, and A.G.Keskar, "Face Recognition Using Contour Matching", IAENG International Journal of Computer Science, Vol: 35, No: 2, 2008.
[15] M. H. C. Law and A. K. Jain, “Incremental nonlinear dimensionality reduction by manifold learning,” IEEE Trans. Pattern Anal. Mach. Intell., Vol. 28, No. 3, pp. 377–391, Mar. 2006.

[16] V.V. Starovoitov, D.I Samal, D.V. Briliuk, "Image Enhancement for Face Recognition", International Conference on Iconics, St.Petersburg, Russia, 2003.
[17] Mittal, N. Walia, E., "Face Recognition Using Improved Fast PCA Algorithm", IEEE Congress on Image and Signal Processing, Vol: 1, pp: 554 - 558, 2008.
[18] A. Hossein Sahoolizadeh, B. Zargham Heidari, and C. Hamid Dehghani, "A New Face Recognition Method using PCA, LDA and Neural Network", International Journal of Computer Science and Engineering, Vol: 2, No :4, 2008.
[19] Deng Cai, Xiaofei He, Jiawei Han, Hong-Jiang Zhang, "Orthogonal Laplacianfaces for Face Recognition", IEEE Transactions On Image Processing, 2006
[20] Ben Niua, Qiang Yang, Simon Chi Keung Shiu, Sankar Kumar Pal, "Two-dimensional Laplacianfaces method for face recognition", Pattern Recognition, Vol: 41, pp: 3237 – 3243, 2008.
[21] "AT&T database of faces" from http://www.cl.cam.ac.uk/ research/dtg/attarchive/ facedatabase.html
[22] M.A.O. Vasilescu, D. Terzopoulos. "Multilinear analysis for facial image recognition," In Proc. Int. Conf. on Pattern Recognition, Quebec City, August 2002.
[23] J.Harguess, J.K.Aggarwal, "An average half face in 2D and 3D Face recognition" in Pattern Recognition and Machine Vision, 2010
[24] Haiping Lu, K. N. Plataniotis and A. N. Venetsanopoulos, "Boosting Discriminant Learners for Gait Recognition Using MPCA Features", EURASIP Journal on Image and Video Processing, 2009, doi:10.1155/2009/713183.
[25] Kela, N. Rattani, A. Gupta, P, "Illumination Invariant Elastic Bunch Graph Matching for Efficient Face Recognition", Computer Vision and Pattern Recognition Workshop, pp: 42, 2006.
[26] Haiping Lu, K.N.Plataniotis, A.N.Venetsanopoulos, "Uncorrelated Multilinear Principal Component Analysis through Successive Variance Maximization" in 25th International Conference on Machine Learning, 2008
[27] M.Kirby, L.Sirovich, "Application of the Karhunen-Loeve Procedure for the characterization of Human Faces", 1990
[28] Yang, J., Zhang, D., Frangi, A. F., & Yang, J. (2004). Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Trans.Pattern Anal. Mach. Intell., 26, 131-137.
[29] Ye, J., Janardan, R., & Li, Q. (2004). GPCA: An efficient dimension reduction scheme for image compression and retrieval. The 10th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (pp. 354{363).
[30] Xu, D., Yan, S., Zhang, L., Zhang, H.-J., Liu, Z., & Shum;, H.-Y. (2005). Concurrent subspaces analysis. Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (pp. 203{208).
[31] Ye, J. (2005). Generalized low rank approximations of matrices. Machine Learning, 61, 167-191.
[32] Wangmeng Zuo, Kuanquan Wang, David Zhang, Hongzhi Zhang, "Combination of two novel LDA-based methods for face recognition", Advanced Neurocomputing Theory and Methodology, Vol: 70, No: 4-6, pp: 735-742 , January 2007.

AUTHORS PROFILE

Shermina.J received the Master degree in Computer Science and Engineering in the year 2002 from Madurai Kamaraj University, India. She began to pursue the Ph.D degree in the Department of Information Technology in Kalasalingam University, India in 2008. Currently she is working as a lecturer in Muscat College affiliated to the University of Stirling, UK. She is a member of IEEE and the current research interest is in the field of image processing.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Performance Analysis of Estimation of Distribution Algorithm and Genetic Algorithm in Zone Routing Protocol

Mst. Farhana Rahman
Computer Science and Engineering Discipline
Khulna University
Khulna 9208, Bangladesh
E-mail: farha_cseku@yahoo.com

S. M. Masud Karim
Computer Science and Engineering Discipline
Khulna University
Khulna 9208, Bangladesh
E-mail: masud@cse.ku.ac.bd

Kazi Shah Nawaz Ripon
Computer Science and Engineering Discipline
Khulna University
Khulna 9208, Bangladesh
E-mail: ripon@cse.ku.ac.bd

Md. Iqbal Hossain Suvo
Computer Science and Engineering Discipline
Khulna University
Khulna 9208, Bangladesh
E-mail: suvo_rgbd@yahoo.com

***Abstract*—In this paper, Estimation of Distribution Algorithm (EDA) is used for Zone Routing Protocol (ZRP) in Mobile Ad-hoc Network (MANET) instead of Genetic Algorithm (GA). It is an evolutionary approach, and used when the network size grows and the search space increases. When the destination is outside the zone, EDA is applied to find the route with minimum cost and time. The implementation of proposed method is compared with Genetic ZRP, i.e., GZRP and the result demonstrates better performance for the proposed method. Since the method provides a set of paths to the destination, it results in load balance to the network. As both EDA and GA use random search method to reach the optimal point, the searching cost reduced significantly, especially when the number of data is large.**

Keywords-Mobile Ad-hoc Network, Zone Routing Protocol, Estimation of Distribution Algorithm, Genetic Algorithm

I. INTRODUCTION

A *Mobile Ad-hoc Network* (MANET) is a collection of mobile nodes that dynamically form a temporary network. It forms the temporary network without any support of infrastructure. So, in the network there are possibilities of lack of reliability and unwanted delay. Again, if the number of nodes grows, the linear search will become costly and the complexity will become high. In case of large number of nodes, a random search will be beneficial where the worst case will equal the linear search. Because of frequently changing topology, low transmission power and asymmetric links routing protocols, MANET have to face the challenge for routing. *Zone Routing Protocol* (ZRP) is a widely used protocol for MANET. In 1997, ZRP was first introduced by Haas [2]. It was proposed to reduce the control overhead of proactive routing protocols and to decrease the latency caused by routing discover in reactive routing protocols.

Recently the scope of Genetic Algorithm (GA) has been extended to solve the ZRP problems. The GA has performed better in the sense of huge search space reduction, while guaranteeing the convergence of the solution. The GA is an adaptive heuristic search algorithm premised on the evolutionary ideas of natural selection and genetics [8]. The basic concept of GA is designed to simulate processes in natural system necessary for evolution. GA represents an intelligent exploitation of a random search within a defined search space to solve a problem. *Estimation of Distribution Algorithms* (EDA) [6], sometimes called Probabilistic Model-Building Genetic Algorithms (PMBGA), are an outgrowth of GA. In a GA, for an optimum solution a population of candidate solutions to a problem is maintained as part of the search. This population is typically represented as an array of objects. Here GA plays an important role in optimizing the search. This is because GA calculates the fitness of each population and generates a better population using crossover and mutation. So, the chance of getting good solutions increases dramatically. But due to the trap of local optima and the widespread diversity of solutions situations may occur where GA never converge to the optimal point that is failed to find a path which is existing between zones. And in some of the cases, GA takes longer time than expected to find a path.

This is the point where EDA works better than GA. Strictly speaking; GA and EDA are same apart from the crossover and mutation. There is nothing called crossover and mutation in EDA. Instead they use probabilistic model for generating new

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

population. This guarantees the better generation of population than earlier generation. Also EDA converges faster, even if there is no feasible routing path. Thus the point is beneficial in the sense of performance of reduction in time and number of generations over GA by EDA. Here we prove the above; that is EDA finds routing path to a destination with minimum time and cost then GA from source when the source and destination is in different zone and number of node involved is large (about 100 to 1000 or more).

As in the mid to late 1990s, laptops and 802.11/Wi-Fi wireless networking became widespread, for research MANET became a popular subject. Many protocols have been proposed for routing in MANET. These protocols can broadly be classified into two types: *proactive* and *reactive* routing protocols. On case of proactive or *table-driven* protocol, by broadcasting routing updates in the network routes to all the nodes is maintain such as *Destination-Sequenced Distance Vector* (DSDV), whereas for reactive or *on-demand* protocols a route to the destination is determined only when the source attempt to send a packet to the destination such as *Dynamic Source Routing* (DSR). Using routing tables, proactive protocols maintain the routing information from one node to the other. Whenever the source has to send any packet to the destination, using the routing tables, path to destination can be found incurring minimum delay. But it may result in a lot of wastage of the network resources if a majority of these available routes are never used. Usually reactive protocols are associated with less control traffic. In a dynamic network a node has to wait until a route is discovered and a route discovery is expensive [7]. Also this causes unnecessary wastage of network resources and also wastage of time [5].

Hybrid protocols combine features of both reactive and proactive routing protocols. The ZRP is a hybrid protocol. It consists of proactive *Intra-zone Routing Protocol* (IARP), reactive *Inter-zone Routing Protocol* (IERP), and the *Border-cast Resolution Protocol* (BRP). ZRP works well both for table-driven protocols and on-demand protocols. But it provides short latency for finding new routes. Decision on the zone radius has significant impact on the performance. In ZRP, the actual problem comes when the destination is outside the zone. In this case, it makes use of Route Discovery with IERP, BRP and uses linear searching on the nodes. This process is time consuming and searching complexity arises as number of node involves increases [5] [7].

In order to detect new neighbor nodes and link failures, the ZRP relies on a Neighbor Discovery Protocol (NDP) provided by the Media Access Control (MAC) layer. NDP [9] transmits "HELLO" beacons at regular intervals. Upon receiving a beacon, the neighbor table is updated. Neighbors, for which no beacon has been received within a specified time, are removed from the table. If the MAC layer does not include a NDP, the functionality must be provided by IARP. Route updates are triggered by NDP, which notifies IARP when the neighbor table is updated. IERP uses the routing table of IARP to respond to route queries. IERP forwards queries with BRP. BRP uses the routing table of IARP to guide route queries away from the query source.

Recently GA has been used in MANET to find the optimized solution [1] [3] [4] [7]. A large amount of work has been done on the application of GA or evolutionary algorithms to communications networks.

Our objective is to use ZRP as an application in EDA, and compare the performance with the method used by the GA.

II. Literature Review

A. Zone Routing Protocol

The ZRP is based on the concept of zones [2]. For all the nodes in the zone, a routing zone is defined separately. The routing zone is based on the radius *r* which is then expressed in hops. Thus, the nodes included in the zone of a node are a maximum of radius *r* away from the node. In Fig. 1, the routing zone of *S* includes all the nodes from *A* to *I* but not *K*, as it resides further than the radius *r*. It should however be noted that the zone is defined in hops, not as a physical distance.

There are two types of nodes in a zone. The nodes residing with an exact distance of radius *r* are the peripheral nodes, and all the other nodes within the circles are interior nodes. The nodes are connected with each other bidirectional, if there is a routing path within the nodes. Intermediate nodes can be used to reach another node, based on the objective function. For example, in Fig. 1, we can reach node *H* from *S* by two possible ways; however only one route is chosen based on the objective criterion. A detail of ZRP can be found in [2] for further reading.

B. Genetic Algorithm

GA [8] is an evolutionary approach to reach to an optimal point in a search space. For larger search space, GA becomes more meaningful and it reduces the searching time, explores in various dimensions within the search space using different GA techniques, like *crossover*, *mutation* etc. Although there are possibilities to trap in the local optima, there are several ways of getting out of it using crossover and thus reach global optima.

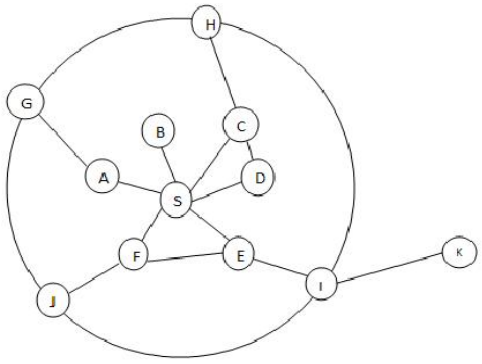


Figure 1. Routing zone of *S* with *r* = 2.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The outline of basic GA with description is given below:

1. **[Start]** Generate random population of *n* chromosomes (suitable solutions for the problem)
2. **[Fitness]** Evaluate the fitness *f(x)* of each chromosome *x* in the population
3. **[New population]** Create a new population by repeating following steps until the new population is complete
   (a) **[Selection]** Select two parent chromosomes from a population according to their fitness (the better fitness, the bigger chance to be selected).
   (b) **[Crossover]** With a crossover probability cross over the parents to form new offspring (children). If no crossover was performed, offspring is the exact copy of parents.
   (c) **[Mutation]** With a mutation probability mutate new offspring at each locus (position in chromosome).
   (d) **[Accepting]** Place new offspring in the new population.
4. **[Replace]** Use new generated population for a further run of the algorithm.
5. **[Test]** If the end condition is satisfied, **stop**, and return the best solution in current population.
6. **[Loop]** Go to step **2**.

As we can see from the GA outline, the *crossover* and *mutation* are the most important parts of the algorithm. The performance is influenced mainly by these two operators. Detail of GAs can be found in [8].

C. Estimation of Distribution Algorithm

In EDAs [6], the problem specific interactions among the variables of individuals are taken into consideration. It is the most recent adaptation of evolutionary approaches. It is starting to be widely used as a promising alternative of GA. The evolving process of EDA is the same as GA apart from crossover and mutation. Instead, EDA uses probabilistic distribution. The probability distribution is calculated from a database of selected individuals of previous generation. The pseudo code of EDA can be formulated as follows:

1. **[Start]** $D_0 \leftarrow$ Generate M individuals (the initial population) at random.
2. **[Fitness]** Evaluate the fitness *f(x)* of each chromosome *x* in the population. Repeat steps 3 to 5 for $l = 1, 2 \ldots$ until the stopping criteria met.
3. **[Selection]** $D_{l-1}^{se} \leftarrow$ Select $N <= M$ individuals from D_{l-1} according to selection method.
4. **[Estimation]** $p_l(X) = p\left(X \middle| D_{l-1}^{se}\right) \leftarrow$ Estimate probability distribution of an individual being among the selected individuals.
5. **[New Population]** $D_l \leftarrow$ Sample M individuals (the new population) from $p_i(x)$.

The easiest way to calculate the estimation of probability distribution is to consider all the variables in a problem as univariate. Then the joint probability distribution becomes the product of the marginal probabilities of *n* variables, i.e.,

$$p_i(x) = \prod_{i-1}^{n} p(x_i). \tag{1}$$

D. Univariate Marginal Distribution Algorithms

In UMDA [6], it is assumed that is there is no interrelation among the variables of the problems. Hence the *n*-dimensional joint probability distribution is factorized as a product of *n* univariate and independent probability distribution. That is:

$$p_i(X) = p\left(X \middle| D_{i-1}^{se}\right) = \prod_{i-1}^{n} p(x_i). \tag{2}$$

The pseudo code for UMDA is as follows:

1. $D_0 \leftarrow$ Generate M individuals (the initial population) at random.
2. Repeat steps 3 to 5 for $l = 1, 2\ldots$ until stopping criteria met.
3. $D_{l-1}^{se} \leftarrow$ Select $N \le M$ individuals from D_{l-1} according to selection method.
4. Estimate the joint probability distribution

$$p_i(X) = p\left(X \middle| D_{i-1}^{se}\right) = \prod_{i-1}^{N} p(x_i). \tag{3}$$

5. $D_l \leftarrow$Sample M individuals (the new population) from $p_l(x)$.

In UMDA the joint probability distribution is factorized as a product of independent univariate marginal distribution, which is estimated from marginal frequencies:

$$p_i(x_i) = \frac{\sum_{j-1}^{N} \delta_j\left(X_i = x_i \middle| D_{i-1}^{se}\right)}{N} \tag{4}$$

with $\delta_j\left(X_i = x_i \middle| D_{i-1}^{se}\right) = 1$, if in the *j*th case of D_{i-1}^{se}, $X_i = x_i$; 0 otherwise.

III. Proposed Method

We choose a random network with maximum chromosome length *N*, and applied ZRP to determine different zones with a radius of *r*, where *r* is the maximum distance of a node from the central node of a zone. This gives us simplified form of a route from the source node to the destination node using border nodes. Using this route as a chromosome, we create the

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

population and apply GA and EDA. The GZRP uses the popular encoding scheme and the minimizing fitness function.

S	B_1	B_2	… …	B_n	D

Where, S = Source, B_i = Border node, D = Destination.

Figure 2. Chromosome formation

Here, all the nodes belong to different zones. The initial population is created randomly, containing the individuals of the above chromosome format. The minimizing fitness function would be the one with finding the shortest route from the source to the destination. The function can be given as follows:

$$I_{ij} = \begin{cases} 1 & \text{if the link from node } i \text{ to node } j \text{ exists} \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Thus, we choose the objective function as the cost of two interconnected nodes multiplied by I_{ij}. In case of GA, we apply one-point crossover and mutation to get rid of stacking in local optima and increase the diversity of solutions. The point is chosen randomly, and crossover is applied between the two randomly selected individuals. Mutation operator then flips the randomly selected genes of the newly formed chromosome with the partial route from the mutation point.

In case of EDA, we use the same encoding scheme and the fitness function. As there is no crossover and mutation in EDA, the only challenge was to compute the probabilistic distribution of chromosome. The problem in this case is trivial. All the chromosome lengths are not the same, from the source node to the destination nodes. So, we apply the technique of continuous EDA domain, where the probabilistic model is generated using the mean and standard deviation. Thus the random function used in this case is the normal distribution function. In both the experiments, we use two terminating criterion, namely, maximum number of round in a single run and the converged solutions. Whenever we reach a converged solution, the program terminates, and if the program cannot converge to an optimal solution, we stop the run after a fixed number of iteration.

IV. Experiment and Result Analysis

In our proposed method, we use the maximum number of iteration in a single run as 1000. The network length varies from 100 to 1000. The sub-population size used in EDA is 50 percent of the main population. The mutation factor is used as 90 percent, meaning a high probability of mutation chance for each individual. As we want to apply our method to ZRP, we do not use any benchmark data set of networks; rather try to handle the situation of dynamically formed network. Thus we increase the network size from 100 to 1000 with the increment of 100 nodes each time. Then we run the program for each set 10 times and used the average of the solution.

Figure 3 shows the performance of GA and EDA in terms of required number of Generations. This figure gives a clear view that, for the above parameter settings, EDA outperforms GA when the network grows in size. For the simplicity in implementation, we considered only the cost of the routing path in a zone.

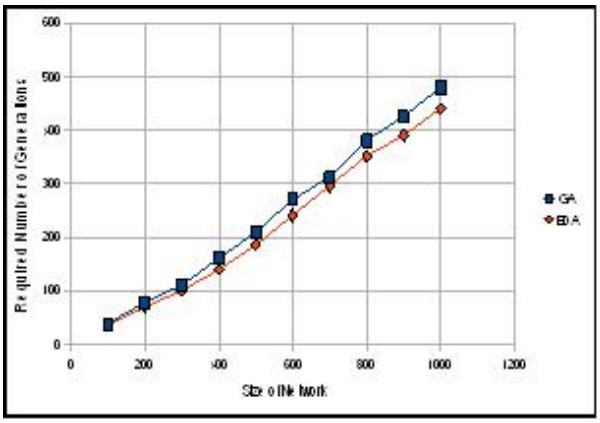


Figure 3. Required number of generation to find the converged value for the same network size using GA and EDA. *(figure caption)*

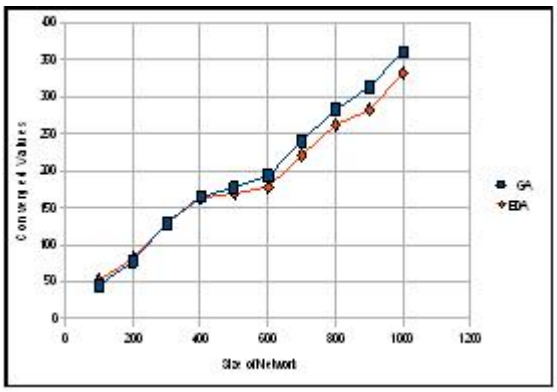


Figure 4. Converged values determined by GA and EDA.

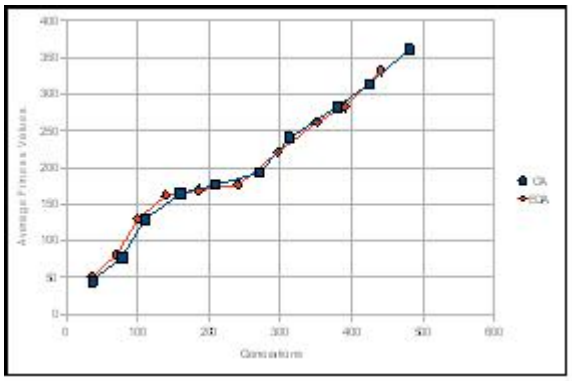


Figure 5. Average fitness values determined by GA and EDA of 50 individual runs.

In Figure 4 the best value of the 50 individual runs is taken to measure the performance against the average number of generations. Here it can be seen that GA performs better when the network size is small. In this case, when the network size grows more than 400 EDA performs better by resulting in a lower converged value for large size network. Thus our approach of applying EDA to solve ZRP proved to perform

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

better than GA. Figure 5 states the average fitness values (in this case the optimal routing path cost) in each generation by EDA and GA.

Again, as our objective function was minimizing, the lower values of EDA indicates better solutions over GA. Also we are obtaining a set of routing path from source to destination from generation. As the network is generated randomly the same routing path is not used as the shortest path from source to destination which results in low traffic and load balance in the network.

V. CONCLUSION

Evolutionary approaches are not guaranteed to find the optimal solutions but they can minimize the cost significantly and are proved effective in larger search space. EDA is a growing field in evolutionary approaches and becoming popular day by day. Our contribution opens a new scope of applying EDA in such a field like ZRP, where GA is already applied. In this study, we only consider UMDA. In future we can extend our work to apply population based incremental learning (PBIL) algorithm [6] and Compact Genetic Algorithm (CGA) [6] both of which are forms of EDA. Then we can decide the best EDA approach to solve ZRP. Again, the path rediscovery can be solved in case of a break down in the network by EDA and GA. Thus, we can again compare the performance in this aspect.

REFERENCES

[1] H. Cheng, J. Cao, X. Fan, "GMZRP: Geography-aided Multicast Zone Routing Protocol in Mobile Ad Hoc Networks", in proceeding of Mobile Networks and Applications Conference 2008.

[2] H. J. Haas, "A new routing protocol for the reconfigurable wireless networks", in proceeding of IEEE 6th International Conference on Universal Personal Communications 97, 1997, pp. 562-566.

[3] J. Inagaki, M. Haseyama, H. Kitajima, "A genetic algorithm for determining multiple routes and its applications", in proceeding of IEEE Int. Symp. Circuits and Systems, 1999, pp. 137-140.

[4] J. M. Kin, T. H. Cho, "Genetic Algorithm Based Routing Method for Efficient Data Transmission in Sensor Networks", in proceeding of ICIC 2007, pp. 273-282.

[5] M. Pelikan, D. Goldberg, F. Lobo, "A Survey of Optimization by Building and Using Probabilistic Models", Illinois: Illinois Genetic Algorithms Laboratory (IlliGAL), University of Illinois at Urbana-Champaign.

[6] D. Turgut, S. Das, R. Elmasri, B. Turgaut, "Optimizing clustering algorithm in mobile ad hoc networks using genetic algorithm approach", in proceeding of IEEE Global Telecommunications Conference, 2002, pp. 62-66.

[7] D. Whitley, "A Genetic Algorithm Tutorial", Statistics and Computing, Vol. 4, No. 2, 1994, pp. 65- 85.

[8] P. Sateesh Kumar, S. Ramachandram, "The performance evaluation of cached Genetic Zone Routing Protocol for MANETs", ICON 2008, pp. 1-5.

[9] C. S. R. Marthy, B. S. Manoj, "Ad Hoc Wireless Networks Architecture & Protocols", ISBN 81-297- 0945-7, Pearson Education Pvt. Ltd, Singapore.

AUTHORS PROFILE

Mst. Farhana Rahman was an undergraduate student of Computer Science and Engineering (CSE) Discipline, Khulna University, Bangladesh. She has started his B.Sc.Engg.(CSE) degree in 2005. She did his undergraduate thesis in the field of evolutionary computing. She has particularly shown his keen interest in zone routing protocol in Mobile Ad-hoc Network.

S. M. Masud Karim has been serving as a faculty member of Computer Science and Engineering (CSE) Discipline, Khulna University, Khulna, Bangladesh. He completed his B.Sc.Engg.(CSE) degree with distinction in 2001. He went abraod for hisher studies in 2006 and was awarded M.Sc. in Media Informatics from Technical University of Aachen (RWTH Aachen), Germany in 2008 and M.Sc. in Informatics from University of Edinburgh, UK in 2009. His areas of interest include information retrieval, data exchange, data integration, computer security.

Kazi Shah Nawaz Ripon has been serving as a faculty member of Computer Science and Engineering (CSE) Discipline, Khulna University, Khulna, Bangladesh. He completed his B.Sc.Engg.(CSE) degree with distinction in 2000. He completed M.Phil in Computer Science from the City University of Hong Kong in 2006. He is currently doing his Ph.D in the University of Oslo, Norway. His areas of interest include multiobject evolutionary alogirthms, genetic algorithm and computer network.

Md. Iqbal Hossain Suvo is an undergraduate student of Computer Science and Engineering (CSE) Discipline, Khulna University, Bangladesh. He has started his B.Sc.Engg.(CSE) degree in 2005. He did his undergraduate thesis in the field of evolutionary computing.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Security risk assessment of Geospatial Weather Information System (GWIS) : An OWASP based approach

K.Ram Mohan Rao
Geoinformatics Division
Indian Institute of Remote Sensing (NRSC)
Dehradun, India
Email: rammohan@iirs.gov.in

Durgesh Pant
Department of Computer Science
Kumaun University
Nainital, India
Email: durgesh.pant@gmail.com

***Abstract*—Security assessment is crucial in web application development environment. The Rapid Application Development (RAD) process makes the application extremely short and makes it difficult to eliminate the vulnerabilities. Here we study how web application risk assessment technique such as risk rating process can be applied to web application. We implement our proposed mechanism the application risk assessment methodology using Open Web Application Security Project (OWASP) model for the security assessment of web application. The study led to quantifying different levels of risk for Geospatial Weather Information System (GWIS) using OWASP model.**

Keywords-Security assessment;Rapid Application Development; Risk rating.

I. INTRODUCTION

Web application security assessment is a crucial part in the application development cycle. The distributed nature of web and application architecture creates difficulties in analyzing the application [1]. The Rapid Application Development (RAD) process makes the application extremely short and makes it difficult to eliminate the vulnerabilities. In the process most of the applications are becoming vulnerable to attack. Although organizations have many traditional precautions in their network such as firewalls, they are no longer sufficient to protect the application across the internet. Firewalls alone can't protect the application from the external threat, but firewalls are integral part of the network security. To withstand web application against the evolving threat techniques, organizations must assess their web applications so that they understand the risk they are dealing with. For example, agency providing data access to their users via the web application must test the web application and calculate or rate the risk. This approach provides, to understand the vulnerabilities associated with the application. Most of the cases this can be achieved by scanning the sites with legacy tools which will detect the number of vulnerabilities present in the site. This will give the opportunity to rectify the coding techniques to eliminate the vulnerabilities.

Knowing the vulnerabilities alone will not help the management to improve the security of the application. Rating the risk of the application by considering the different factors associated with application will give more clarity and edge to secure the application in a better way. By following this approach, organization can estimate the severity of the application and make an informed decision about the risk. Also the risk factors will priorities the issues in the application in a better way than the random approach. The areas having more risk can be immediately looked into, than the next prioritized zone. In this paper, a modest attempt has been made to implement Web Assessment Methodology (WAM) to the Geospatial Weather Information System (GWIS) application and risk factor has been derived by using Open Web Application Security Project (OWASP) risk rating methodology.

II. APPLICATION TESTING FRAMEWORK

Software testing has gone evolutionary process. The global software testing market value is $13 billion [2]. Generally software development is up to 40% of a typical software product release budget, but testing of the software is about 40% of the development budget. Hence testing phase is very important for any software or application for resulting to higher quality software. Software testing is the process used to identify the correctness, completeness, security, and quality of the developed software programs. It is the process of the technical investigation to reveal the quality related information about the product or application. Testing furnishes criticism of the application for further improvement of the application in several contexts. There are three fundamental approaches for the automotive tests of web application [3]. Black box, white box and gray box provide the different approaches for assessing the security of web applications. White box and black box are the terms used to describe the point of view a test engineer takes when designing the test cases. Black box takes the external view of the application and white box takes internal view. That is the application is tested from the inside using its internal application programmatic interface (generally the API) in the white box testing, and the application is tested using its outward facing interface (generally GUIs) with the black box testing. Gray box testing

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

is combination of both white box and black box testing. It allows security analysts to run automated and manual penetration test against a target application. Table 1 shows a comparative statement on testing technologies.

TABLE I. TESTING TECHNOLOGIES: A COMPARISONS

Pros	**Cons**
Manual: Penetration or security acceptance by small set of people using known tools and scripts	
Generates well targeted tests for specific application functions.	1. Limits testing to experts which may lead to bottlenecks. 2. Can lead to a high error rate with recurring costs 3. Limits application coverage due to time constraints.
Automated: Specific tests for individual function, built by the code developer. Quality assessment teams build tests from end user perspective.	
Offsets expenses with improvements in quality, reduced effort for acceptance and iterative development processes.	Requires greater overhead to create and maintain than manual testing.
Black box or System: Looks only at system input and output, modifying normal user input to make the application behave in unintentional ways.	
Uses established automated test tools that require minimal application knowledge to use.	Possible only when application components are ready for testing.
White box or Source: Assesses individual components for specific functional errors, often in combination with code scanning tools and peer reviews.	
Uses tools that have established integrations with developer IDEs, enabling the well-defined discovery of flaws in tested functions.	Does not uncover requirement and design flaws. May not uncover vulnerabilities to attacks involving multiple components or specific timing not covered by unit testing.
Gray box (Using application defined framework): Combines black- and white-box testing to create tests unavailable via commercial tools.	
Provides the most comprehensive method by combining system and unit level testing.	Requires that a framework be specified during the inception phase and design activities. Require much effort to build the test framework to build the application.

Selection of particular testing methodology depends on number of factor(s), such as time allotted to the assessment, access to the internal application resources and goals of the test [3].

Application security is the use of software, hardware and procedural methods to protect applications from external threats. Security measures built into application and sound application security procedures minimize the likelihood of the attack. Security is becoming an increasingly important concern during development as applications become more frequently accessible over network. As a result applications are vulnerable to a wide variety of threats. Application security can be enhanced by rigorously defining enterprise assets, identifying what each application does with respect to these assets, creating security profile for each application, identifying and prioritizing potential threats, and documenting adverse events and the actions taken in each case. This process is known as threat modeling [4].

III. GEOSPATIAL WEATHER INFORMATION SYSTEM

Geospatial Weather Information System (GWIS) is a web based tool for capturing, storing, retrieving and visualization of the weather climatic data. The GWIS contains historical climatic data for nearly hundreds of land stations country wide. The database is provided with both climatic daily and monthly data. Daily data has been nearly for 150 ground stations country wide and covering temperature, rainfall, humidity details. The climatic monthly data has for wide range of land stations around 3000 countrywide. Daily data is being captured from different sources after then arranged in GWIS format for storing in the database. The source for monthly data is Global Historical Climatology Network (GHCN). It is used operationally by National Climatic Data Centre (NCDC) to monitor long-term trends in temperature and precipitation. The mission of GWIS is to integrate the weather related information from different available sources and organize the data in structured GWIS format. The application tool is designed to cater the research needs of various application scientists working on different themes.

IV. RISK ASSESSMENT METHODOLOGY

Performing web application assessment is a difficult task because of its complex application architecture. The task should be like any other software testing process – with a methodology, testing procedures, a set of helpful tools, skills and knowledge [5]. In general, risk model involves several factors such as asset information based on their importance in business, likely threats to these assets, associated vulnerabilities (both technical and non-technical), severity levels of the vulnerability, business impact factors such as company reputation. These factors weightage depends on the organization structure, its goals, the impact on its application business etc. Some models may give more weightage to technical factors and some may give weightage to financial

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

factors, but overall it depends on the strategy of the organization and its goals.

There are number of risk assessment models namely CVSS, OWASP, CENZIC, AS/NZ, OCTAVE, NIST, NISA, ISO 1799 and ISO 27001 for assessing the risk associated with the application. In addition to these models there were number of other testing techniques available for web application testing [6, 7, 8], analysis [9], and reverse engineering [10, 11]. Huang et al,2003 [12] proposed a WAVES (Web Application Vulnerability and Error Scanner) – black box testing framework for automated web application security assessment. Scott and Sharp, 2002 used an application proxy to abstract Web application protection strategy [13]. Huang et al, 2004 describe the bounded model checking (BMC) for verifying the web application code for automatic patching of vulnerable code with run time guards allowing both verification and assurance occur without users intervention [14]. Again Huang, 2004 proposed a holistic approach to ensure the web application security by static analysis and runtime protection [15]. In addition to these some more white box techniques that protect the web application at the development time rather than the deployment phase [16]. In all these listed models and techniques the general assessment framework involves:

- Identifying assets, possible threats and vulnerabilities.
- Estimating the risk.
- Determining the severity of risk.
- Deciding the priority list to eliminate the vulnerabilities.
- Finally, Customizing risk model with given inputs.

The ultimate objective of any model is to quantify the risk of the application at different levels by taking several factors such as asset information, likely threats, associated vulnerabilities, impact of the application. It also provides useful mechanism for organizations to prioritize their business risks with common remedial practices for effective security practices. Risk assessment process involves steps in identifying asset information, possible threat and vulnerabilities of the application. The process ultimately determines the overall risk factor involved with the application and will help the organization to eliminate the vulnerabilities of the application.

A. Identifying assets, possible threats and countermeasures

The first step for an organization to assess the network for security vulnerabilities is to understand the assets that make up the network. This step, known as discovery, involves identifying all of the servers, workstations, devices, services, and applications running on the network.

B. Assest classification

Asset classification starts with identification of assets of organization. OCTAVE-S model illustrates asset identification as “identification of business process assets”. Most of the cases it is always better to talk to the people who can better understand the organization policy and goals to classify the asset category. Asset is information, capability, an advantage, a feature, a financial or a technical resource that should be defended from any damages, loss or disruption. The damage to an asset may affect the normal functionality of the system as well as the individuals or organizations involved with the systems. Normally in the web application technology assets are database server, application server, and web server.

C. Threat Classification

Threat is a specific scenario or a sequence of actions that exploits a set of vulnerabilities and may cause damage to one or more of the system’s assets [17]. When vulnerability is identified, it requires at least one of the attack technique to hack the application. Figure 1 shows the common threat areas, security concerns of application. However specific threats to individual application may differ from one application to other.

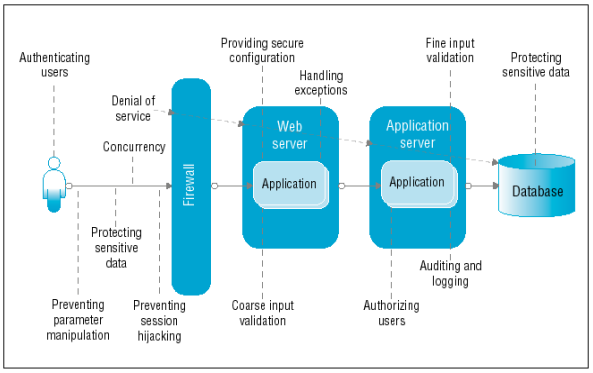


Figure 1. Web application security concern areas [18]

There are several attack techniques available and documented commonly referred as classes of attack. Web Application Security Consortium (WASC), 2004 created and documented threat classification with detailed information about the possible classes of attacks (table 2) which is quite useful information for application developers, security auditors, professionals and vendors [19].

TABLE II. WASC THREAT CLASSES

Threat Classes	**Description**	**Attack**
Authentication	Checks the identity of user and/or service of application	Brute Force Insufficient authentication
Authorization	Checks the user and/or service permission to perform a request action	Credential session prediction Insufficient authorization
Client side attacks	Checks for a chance to abuse or	Content spoofing, Cross site scripting

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

	exploitation of web site user	
Command execution	Exploit the vulnerability through remote command methods.	Buffer overflow LDAP injection OS commanding SQL injection SSI injection Xpath injection
Information disclosure	technique to acquire system specific information such as software , version, patch level etc.	Directory indexing Information leakage Path traversal Predictable resource location.

D. Finding out vulnerabilities

Because of the multiple technologies involved in the web application, it's very difficult to discover the vulnerabilities present in the site. For this, tester has to run battery of tests and use various strategic techniques to discover the vulnerabilities. This process includes,

- Discovering any existing vulnerabilities,
- Verify the discoveries,
- Document the verified discoveries in a standard template.

Manual testing is a cumbersome method and almost a difficult task for discovering the application flaws. Discovering vulnerabilities present in the application often needs a strong testing tool for scanning the web site. Recently, automated security testing tools are developed for scanning the total application with easy to use interface. Such projects are WebSSARI (Web Application Security Analysis and Runtime Inspection) for testing vulnerabilities in PHP code [20]. However, such white box testing techniques fail to adequately consider the runtime behavior of web applications [21]. The main reason for this is web application is a combination of presentation layer, application layer and database layer. In other words, white box testing techniques will not give appreciated results to discover the vulnerabilities because their inability in handling user input through the interfaces (GUIs). There is other approach for finding vulnerabilities of web application, a black box testing technique. Huang et al, (2005) presented a web application security scanner (WAVES)- an automated software testing platform for the remote, black box testing [22]. Barton et al, (2000) presented a web application security assessment framework based on a security scoring vector [23].

Today there are several automated tools, works on a black box testing framework that automates the process of web application security testing and finally result the vulnerabilities present in the application. The tools analyses any web applications or sites and scans them for exploitable vulnerabilities such as Blind SQL Injection, Cross site scripting, and Directory traversal, etc. Lot of commercial and open source tools are available for the application assessment. There are several free and open source projects in the category of web application assessment tools, such as Burp proxy, Grabber, Penetra, Paros, SPIKE Proxy, WebScarab, Wapiti, NESSUS and W3AF. Commercial assessment tools includes the list of AppScan, Webinspect, BiDiBLAH, Cenzic, Acunetix , , Wikto, Nikto, etc. In general all these automated tools generates following feature set at high level:

- Vulnerability discovery and listing
- Automated auditing as well as manual auditing capabilities
- Extensive view options
- Extensive reporting options
- Exposed APIs
- Remediation recommendations

Identifying vulnerabilities across the site is a major endeavor. Today's enterprise consists of several system servers, application servers, database servers via several networking circuits with varying speeds. The point is, it is not possible to simply install network or system scanners and scan the total application. This is because, it is not possible to get the required coverage with in the desired time frame with a single scanner. For this reason we cannot simply stop the assessment, knowing that the site consists of 70 percent network vulnerabilities that have not been remediated. Enterprise level assessments are still required. Instead of simply dropping scanners onto network, the process should leverage organizations vulnerability management, its investment in security, patch, and configuration management technologies. Vulnerability scanners are responsible for detecting network hosts, discovering available applications, and ascertaining vulnerabilities. Vulnerability softwares generally run on network devices or on a company own application assets. For this type of application assessment, single type of vulnerability scanner is sufficient for scanning the application.

TABLE III. VULNERABILITIES BY PATTERNS

No	Vulnerability Patterns	Number of Instances
1	Blind SQL Injection	2
2	Login page SQL Injection	2
3	Unencrypted login request	1
4	Application Error	1
5	Inadequate account lockout	1
6	Permanent cookie contains sensitive session information	1
7	Session information not updated	1
8	Unencrypted password parameter	3
9	Unencrypted viewstate parameter	7

However, larger sites may require multiple vulnerability scanners to support the assessment needs. The reason is the specific tools are effective in some of the areas and may not be good at other functional areas. For this reason, the GWIS

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

application has been scanned with multiple scanners namely AppScan, CENZIC, and Nessus tools. The consolidated list of vulnerabilities observed is shown in table 3. Discovering the vulnerabilities after a systematic scanning as mentioned above, verification has to be done with proper documentation using a standard template. Figure 2 shows the vulnerability documentation using AppScan template for GWIS. The blind SQL injection is categorized in to high sever vulnerability, occurs in application level. This vulnerability is dangerous in exploiting the application to view, modify or delete database entries. This kind of detailed template helps the developer and designer to address the vulnerability immediately with the remediation procedures.

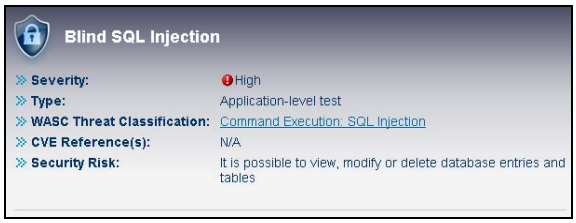


Figure 2. Vulnerability documentation template

V. Assessment Models

There are many assessment models for rating risk associated with the applications. Each one of them has their own merits and differs by what they measure. But there is no standardization of web application security assessment process. One of the issues with the assessment is, there is no standard metric to measure the application security. For example OWASP measures the application security by categorizing the risk in the scale of 1-10, and divide the scale in 3 bands, that is low, medium and high [24].

The (OWASP) covers many projects over application security aspects, such as documentation, tools, models etc. Open Web Application Security Project (OWASP) risk model quantifies risk in the scale of low, medium and high. The model rates the risk by taking input parameters such as attacker capability, vulnerabilities of the application and impact of a successful exploit on the business. The Common Vulnerability Scoring System (CVSS) provides similar risk rating approach to communicate characteristics and impacts of application vulnerabilities. CVSS works on Base, Temporal, and Environmental metrics for the application assessment [25]. The Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE) defines risk based strategy assessment and planning technique for the security [26]. Cenzic also provides the application security scored called as Hailstorm Application Risk Metric (HARM) based up on quantitative score of risk quantified as a numeric number. It also considers the base risk, impact areas of the application, risk factors such as attack complexity, attack boundaries, detection precision, and asset weights in quantifying the risk. For a given application the HARM score is calculated using a series of proprietary equations [27] (Cenzic, 2008). Huang et al. proposed a WAVES (Web Application Vulnerability and Error Scanner) – black box testing framework for automated web application security assessment [28]. But it is true that no model will fit into everyone's need perfectly. The choice of model purely depends on the suitability of the model, business aspects, the technical aspects of application, and the type of metric etc. The study follows the integrated OWASP approach for estimating and rating the risk factor, because of the model suits to the requirements of GWIS environment.

VI. Integrated Owasp Risk Assessment Methodology

OWASP risk rating methodology is a simple approach for calculating and rating the risk associated with the application. It is an integrated approach of several existing models for representing the overall severity of the application. The system is not entirely quantitative rather the system is formulated, to rely on a base qualitative ranking of risk into number of categories.

In a nut shell, this risk rating calculation is the process of gathering details about the threat agent involved, attacker ability of skills, motive, the opportunity, the size of attacker domain, and vulnerability factors associated with the application, such as ease of discovery of the vulnerabilities, ease of exploit, awareness of the vulnerability and finally intrusion detection. One critical observation in this model is severity of the vulnerability which is most important parameter in estimating the risk is missing. In the general sense, severity of the vulnerability of the application carries more weightage and dictates the risk associated with the application. Hence another factor called severity of the vulnerability is included in the vulnerability factors. Therefore,

Risk = Likelihood * Impact

Where
Likelihood = Chances of vulnerabilities for being exploited by the attacker.
Impact = Impact of successful attack.

There are number of factors associated to calculate the likelihood value. They are threat agent factor and vulnerability factors. Further these factors are divided into the sub factors for calculating the threat agent factors and vulnerability factors.

A. Likelihood calculation

Likelihood is a rough estimate of how likely particular vulnerability is uncovered and the same is exploited by the attacker. There are two major components in estimating the chances for a successful attack. The first one is the capability of threat agent factor and second one is vulnerability factor. Threat agent factor deals with the capability of the group, their skill level, motivation and size of the group.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

There are two major components in estimating the chances for a successful attack. Maximum the capability levels of the group means, maximum are the chances for exploiting the vulnerabilities in the application. Vulnerability factor consider the variety of factors related to the vulnerability presence in the application. Once the application having the vulnerabilities means, there are more chances of exploiting the application depending on factors such as ease of exploitation, awareness of the particular vulnerability, detecting the same with mitigation techniques etc.

B. Threat agent factors

The objective of threat agent factor is to estimate the likelihood of successful attack. The general threat agent factors are skill levels of the attacker group, the motivation of group to succeed in the attack, the resources of attacker group, generally the infrastructure and facilities of the group, and the size of the group.

C. Vulnerability factors

This section deals with the finding out the vulnerabilities present in the application and rating attributes related to the vulnerabilities of the GWIS. But the most tedious part in risk assessment is, finding out the vulnerabilities present in the application. Vulnerabilities are generally flaws in the applications. These can be best discovered by deploying automated legacy testing tools on the application. For example, GWIS is scanned with Appsec automated tool that crawls the application for vulnerabilities discovery. OWASP risk rating methodology considers ease of discovery, ease of exploiting, awareness and intrusion detection as vulnerability factors in estimating the likeliness of uncovering the vulnerability that may be exploited by the attacker. However, an important factor missing in the OWASP approach is the severity factor of the vulnerability present in the application. Hence in the revised model, severity factor is included in the vulnerability factors.

TABLE IV. LIKELIHOOD FACTORS OF GWS

	Threat agent factors				Vulnerability factors				
Name of Vulnerability	Skill Level	Opportunity	Motive	Size	Ease of discovery	Ease of Exploit	Aware-ness	Intrusion Detection	Severity
Blind SQL Injection	9	4	9	6	3	9	3	8	9
Login page SQL Injection	9	4	9	6	3	9	3	8	9
Unencrypted login request	6	4	4	2	2	9	4	1	6
Application Error	3	9	4	2	3	9	2	1	3
Inadequate account lockout	3	9	4	2	4	9	6	8	3
Permanent cookie contains sensitive session information	3	4	4	6	4	9	4	3	3
Session information not updated	6	4	4	6	3	9	3	1	3
Unencrypted password parameter	6	4	4	6	6	9	5	1	3
Unencrypted view state parameter	4	4	1	6	7	9	8	1	3

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

TABLE V. IMPACT FACTORS OF GWS

	Technical impact factors				Business impact factors			
Name of Vulnerability	Loss of Confidentiality	Loss of integrity	Loss of availability	Loss of accountability	Financial Damage	Reputation Damage	Non Compliance	Privacy Violation
Blind SQL Injection	9	7	5	9	1	5	5	5
Login page SQL Injection	9	7	5	9	1	5	5	5
Unencrypted login request	6	5	1	7	3	4	2	5
Application Error	2	3	1	7	3	4	2	3
Inadequate account lockout	2	3	1	7	3	4	2	3
Permanent cookie contains sensitive session information	2	3	1	7	3	4	2	3
Session information not updated	2	3	1	7	3	1	2	3
Unencrypted password parameter	2	1	1	7	3	1	2	3
Unencrypted view state parameter	1	1	1	1	1	1	2	3

There is a proof of concept for the SQL injection, the vulnerability requires more skill levels for the exploitation. Similarly, ease of exploitation and severity levels also maximum for SQL injection. Hence the skill level to exploit the SQL injection score is 9.The attacker will learn the structure of the SQL query, and then use this knowledge to thwart the query by injecting data that changes the query syntax into performing differently than indented. Therefore ease of exploitation is maximum. The remaining factors scores of threat agent and vulnerability factors are shown in table 4.

D. Estimating impact

Impact calculation is about estimating the impact of successful attack to GWIS application. It is a measure of vulnerabilities associated technical factors present in the application and the impact of the application, if the vulnerabilities are exploited by the attacker. There are several factors for estimating the impact feature. OWASP considers technical impact and business impact for deriving the impact of the application.

The scoring system does not consider more than one vulnerability, if the application has more than one number of similar types of vulnerability. For example, GWIS consists of two instances of blind SQL injections and seven unencrypted view state parameters. But finally the scoring has given only for one blind SQL injection, and one unencrypted view state parameter, as the type of vulnerability is same. But when particular type of vulnerability is addressed, total number of instances is taken care. This is because of the reason that, each vulnerability provides equal chance of opportunity for exploiting the application.

E. Technical impact

The primary objective of technical impact is to calculate the magnitude of impact if the vulnerabilities are exploited from the application. The technical impact factors are further divided into four classes namely confidentiality, integrity, availability and accountability. The objective of information security systems is to protect confidentiality, integrity, availability. Thus technical impact factors plays major role in the application risk assessment. The technical impact is the estimate of sum of these technical factors by giving suitable weightages to the individual factors.

F. Business impact

The ultimate objective of risk assessment is to quantify the impact of business, if the vulnerabilities are being exploited by the attacker.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Sometimes it is also known as business impact which plays major role in calculating the overall risk rating process. For example, GWIS is an application indented to disseminate the weather related data to the given users through the application. Here the main business assets are weather repository, related financial transactions and meta data sets. OWASP consider four important factors to estimate the business impact of the application, namely financial damage, damage of reputation, non-compliance and privacy violation of the application, if the application is being hacked. Financial damage is minimal from the business point of view, but reputation damage is more because of the organizational aspect. The reputation damage also affects the further business of the GWIS. The non-compliance and privacy violation of the application is medium. The impact scores for each of the vulnerabilities present within the application are added together, resulting in a total for the application score. This process of adding together vulnerability scores must be sensitive because of different category of scores. Therefore calculating the GWIS score is the act of summing up the score for each vulnerability and corresponding factors. Table 5 shows the impact factors of GWIS. Finally risk is calculated by multiplying the likelihood with impact values, which is derived in the process. The resultant value is then categorized with the table to rate the severity of the risk as given in table 6. Likelihood and impact is estimated to calculate an overall severity for the risk. The final risk values are categorized in to low, medium, and high in the scale of 0 to 9 (table 6).

TABLE VI. RISK CATEGORIZATION VALUES

Likelihood and Impact levels	
0 to < 3	Low
3 to < 6	Medium
6 to 9	High

VII. EXPERIMENTAL RESULTS

The GWIS application has been scanned thoroughly for the vulnerabilities across the presentation, business,
and database layers of GWIS. Nine vulnerability patterns are found including total 20 instances. The likelihood and impact scores are calculated against each vulnerability of the application, and the final scores are derived as shown in table 7 to 10.

Table VII. Threat agent factors of GWIS application

Skill Level	Opportunity	Motive	Size	Total	Risk
5.44	5.11	4.77	4.66	20	5

Table VIII. Vulnerability factors of GWIS

Ease of discovery	Ease of Exploit	Awareness	Intrusion Detection	Severity	Total	Risk
3.88	9.00	4.22	3.55	4.66	25.33	5.06

Table IX. Technical impact factors of GWIS

Loss of confidentia	Loss of integrity	Loss of availability	Loss of	Total	Risk
3.88	3.66	1.88	6.77	16.22	4.05

Table X. Business impact factors of GWIS

Financial Damage	Reputation ion Damage	Non Compliance	Privacy Violation	Total	Risk
2.33	3.22	2.66	3.66	11.88	2.97

The overall score of likelihood and impact of GWIS are 5.0333 and 3.51385 respectively (table 10).

TABLE XI. LIKELIHOOD AND IMPACT SCORES

	Score	Risk category
Likelihood	5.0333	Medium
Impact	3.51385	Medium

VIII. RESULTS AND DISCUSSIONS

The OWASP model provides the open framework for security accuracy assessment of web application security. In order to experiment with OWASP open source model, the study has been chosen the GWIS application to implement security assessment. During the assessment phase, the application flaws are completely assessed with variety of tools for finding out vulnerabilities of the application. The found vulnerabilities are billed with threat agent factors, vulnerability factors to find out the likelihood and impact levels on technical and business functions. Now it is combined them together to get final severity risk rating for the GWIS. Final severity risk levels are obtained from overall risk

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

severity matrix (table 11) by inputting likelihood and impact levels of GWIS.

Table XII. Overall risk severity matrix of OWASP

Overall Risk Severity				
Impact	HIGH	Medium	High	Critical
	MEDIUM	Low	Medium	High
	LOW	Note	Low	Medium
		LOW	MEDIUM	HIGH
	Likelihood			

The integrated OWASP model gives an innovative approach for the study of security assessment of GWIS application. The study has produced excellent results in assessing the security of GWIS application by rating the risk factor associated with GWIS.

It has produced two main factors, one is the likelihood and other is impact factor. It is clearly noted that, the likelihood and impact factors are not static ones and will change from time to time depending on the application design principles and business rules. That is why technical factors and business factors are completely different. But understanding the business context of the vulnerabilities is crucial in assessing the risk of GWIS. Considering threat agent factors, vulnerability factors, technical impact factors, and business impact factors of GWIS application, the scores are varying between 0 to 9 depending on the values of threat agent capability, vulnerabilities, technical impact, and business impact on the GWIS. The objective of deploying OWASP model is to quantify the risk face by the GWIS. The vulnerability score helps the GWIS to better understand the application severity, progress towards security goals such as protecting assets, rectifying the risk levels.

maximum risk against threat agent factor, vulnerability factor, technical impact and business impact. The final outcome of overall risk also because of the severity levels of these two vulnerabilities against both likelihood and impact factors of GWIS. Since the vulnerabilities can be exploited remotely, the ease of exploit is relatively high, and hence it has been given the high rating for the ease of exploit with corresponding severity levels.

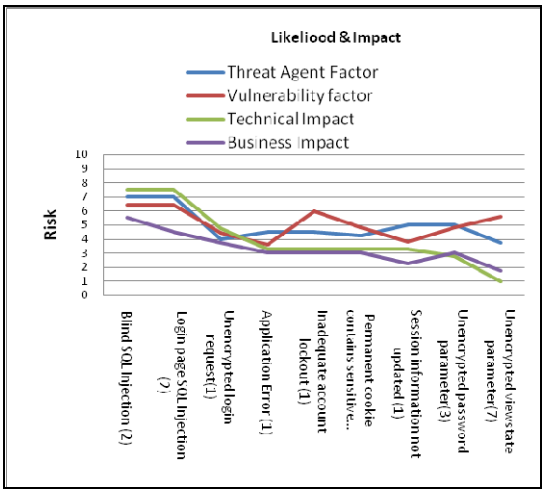


Figure 3: Risk factors of GWIS

Also the vulnerabilities can be exploited using multiple methods with different outcomes scores. To execute the high risk value vulnerabilities such as, blind SQL injection and login page SQL injections, technical and business impact factors are set to HIGH.

TABLE XIII. VULBERABILITIES BY PATTERN & PRIORITY

Vulnerability Patterns	Risk	Resource Requirement	Effort to Exploit	Number of Instances
Blind SQL Injection	High	Low	Medium	2
Login page SQL Injection	High	High	Medium	2
Unencrypted login request	Medium	Low	High	1
Application Error	Low	High	Low	1
Inadequate account lockout	Low	Low	Medium	1
Permanent cookie contains sensitive session information	Low	High	Low	1
Session information not updated	Low	High	Low	1
Unencrypted password parameter	Low	High	Low	3
Unencrypted viewstate parameter	Informational	High	Medium	7

To explain the different categories of risk generating factors of GWIS application, blind SQL injection and login page SQL injection are causing major risk to the GWIS. Figure 3 shows blind SQL injection and login page SQL injection are resulting

Similarly, to exploit the rest of the vulnerabilities the likelihood and impact factors are given in table 3 and 4. As shown in table 10 likelihood of GWIS is MEDIUM, and the technical impact is MEDIUM and hence severity level is

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

medium (table 11). Similarly likelihood is MEDIUM and business impact is LOW, so from pure business point of view, the risk factor is LOW. But on the whole, likelihood and impact levels of GWIS are MEDIUM and MEDIUM respectively. Therefore the overall severity of the risk is MEDIUM. To minimize the risk levels of the GWIS, it is crucial to fix the most sever risk generating vulnerabilities first such as, blind SQL injection and login page SQL injection vulnerabilities in the GWIS. Similarly the other vulnerabilities also should be fixed to further reduce the risk of GWIS in as per the priority list given in table 13. In figure 4 the total number vulnerabilities found in GWIS are depicted severity wise. Out of 20 vulnerabilities, 4 vulnerabilities are high severing levels, 1 is of medium severity, 8 are of low severe levels and rest of them is informational.

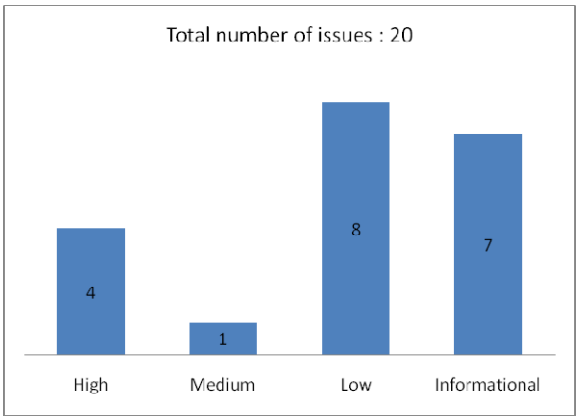


Figure 4: Issue severity gauze for GWIS

Twenty unique issues are detected in GWIS, across 10 sections of the regulations. Table 14 shows compliance scan results related issues of GWIS application according to OWASP threat class categorization.

TABLE XIV. COMPLIANCE SCAN RESULTS

Section	No. of Issues
Cross site scripting (XSS) flaws	-
Injection flaws	**4**
Malicious file execution	-
Insecure direct object reference	**1**
Cross site request forgery (CSRF)	-
Information leakage and improper error handling	**9**
Broken authentication and session management	**6**
Insecure cryptographic storage	**4**
Insecure communications	**4**
Failure to restrict URL access	-

The remediation task is designed to address the vulnerabilities present in the GWIS application. The remediation tasks generally address the weakness of application that are found during the assessment of application. There are different types of vulnerabilities, some requires immediate remediation and some may require some software / hardware resource to rectify the same. Existing code should be checked for these vulnerabilities, as these flaws are being actively targeted by attackers.

Development projects should address these vulnerabilities in their requirements documents and design, build and test their applications. Project managers should include time and budget for application security activities including developer training, application security policy development, security mechanism design and development, penetration testing, and security code review. But all vulnerabilities pose some risk to the application that could result in a loss of system control by compromising the valuable database. The research framework entails that, there are various types of threats a distributed application faces. Those covers information disclosure, infrastructure vulnerabilities, session management flaws, insecure configuration, authorization flaws, encryption flaws, invalidated input, SQL injection, corss site scripting, HTTP response splitting, LDAP injection, and XPath injection. As more and more applications are web based, web application security exploits are becoming the attack patterns for hackers. Exploits embedded in http or https packets sail past perimeter security systems and potentially attack an organization's critical databases. Given the complexity of today's web applications, these exploits are difficult to uncover and protect against. Countermeasures using a combination of open-source tools, automated scanners, and manual testing enumerate vulnerabilities across all the threat class domains. Risk assessment helps companies to proactively deal with security by providing structure and rationale for the security of web application. Threat models help capture security flaws at an early stage, thereby reducing the cost of fixing the flaws after the application has been deployed. Despite the advantages, a number of security challenges to implementing three-tier architecture exist. A full 3-tier implementation would have a server running a web server that connects to a mid-tier server or other servlet engines and database connectors. This arrangement will have access to all the layers of application. Application security is the use of software, hardware, and procedural methods to protect applications from external / internal threats. Security measures built in to applications and sound application security procedures minimize the likelihood to manipulate application. Actions taken to ensure application security are sometimes called as countermeasures. These include firewall, router, encryption/decryption, anti-virus programs, spyware detection/removal programs, and biometric authentication systems.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Application security can be enhanced by rigorously defining enterprise assets, identifying what each application does with respect to these assets, creating a security profile for application, identifying and prioritizing potential threats, and documenting adverse events and the action taken in each case. This process is called as threat risk modeling. The results of the research provide the security assessment of GWIS application, identifying weaknesses and recommending security improvements. In the process, the study presented threat risk modeling, standard vulnerability management by calculating the risk associated with each layer with common remediation initiatives. The study led to quantifying different levels of risk for GWIS by using OWASP approach.

REFERENCES

[1] WN Mills, L Krueger, W Chiu, N Halim, JL Hellerstein, MS Squillante, 2007, Metrics for performance turning of web based application, CMG journal the association of system performance professionals.Vol 12. No.7

[2] Logigear Corporation, 2008, Software testing: The continuous evolution of software testing", Logigear Whitepaper. Accessed from http://www.logigear.com. Accessed on 23.12.2007.

[3] D Cornell, 2007, Web application testing: The difference between white, black and gray box testing", Whitepaper.http://searchsoftwarequality.techtarget.com/tip/0,289483,sid92_gci1246398,00.htm. Accessed on 23.08.2007.

[4] JD Meier , A Mackman, S Vasireddy, M Dunner, S Escamilla, A Murukan, 2003, Improving web application security : Threats and Countermeasures. Microsoft Corporation, pp 45-49.

[5] O Segal, 2006, Methodologies & tools for web application security assessment, Watchfire Whitepaper. Accessed from http://www.blog.watchfire.com, Accessed on 22.03.2008.

[6] N Benedikt, J Freire, P Godefroid, 2002, Veriweb: Automatically testing dynamic websites, in:Proceedings of the 11th International Conference on the World Wide Web, Honolulu, Hawai.

[7] R Gold, 2008,HttpUnit Available from http://httpunit.sourceforge.net. Accessed on 02.05.2008.

[8] F Ricca, P Tonella, ID Baxter, 2001, Restructuring Web applications, in:Proceedings of the 23rd IEEE International Conference on Software Engineering, Toronto, Ontario, Canada.

[9] F Ricca, P Tonella, 2000, Web site analysis : Structue and evolution, in:Proceedings of the IEEE International conference on Software Maintenance, San jose, California.

[10] D Lucca, GAD Penta, M Antoniol, G casazza, 2001, An approach for reverse engineering of web based applications in: proceedings of the Eighth Working conference on Reverse Engineering , Stuttgart, Germany.

[11] S Tilley, S Huang,2001, Evaluating the reverse engineering capabilities of Web tools for understanding site content and structure: A case study in:Proceddings of the 23rd IEEE International conference on Software Engineering, Toronto, Ontario, Canada.

[12] YW Haung, CH Tsai, TP Lin, SK Huang, DT Lee, SY Kuo, 2005, , A testing framework for web application security assessment, computer Networks 48(2005), pp.739-761.

[13] D Scott, R Sharp,2002, Abstracting Application-Level web Security, In:Proceedings of 11th International conference, World Wide Web, May 17-22,2002. (62).

[14] YW Huang, Y Fang, C Hang, CH Tsai, DT Lee, SY Kuo, "Securing web application code by Static Analysis and Runtime protection", in : Proceedings of the 13th International World Wide Web Conference, New York, May 17-22, 2004.

[15] YW Huang, Y Fang, C Hang, CH Tsai, DT Lee, SK Yen, 2004, "Securing web application code by Static Analysis and Runtime protection", Computer Networks 48(2004) pp.788-795

[16] M Bobbit, 2002, Bulletproof Web Security. Network security magazine, TechTarget Storage Media, May 2002. Available from http://infosecuritymang.techtarget.com/2002/may/bulletproof.shtml. Accessed on 22.03.2004.

[17] G Ygor.2005. Practical Threat Analysis for the Software Industry. Software industry.com Whitepaper Available from http://www.securitydocs.com/library/2848. Accessed on 25.04.2007.

[18] A Danny, 2008, Web application security : Automated scanning versus manual penetration testing. Rational Software, IBM Software group. Whitepaper. Accessed from http://www.ibm.com. Accessed on 12.07.2006. WASC, 2004, Web Application Security Consortium: Threat Classification, Whitepaper. Available from http://www.webappsec.org/projects/threat. Accessed on 03.03.2006.

[19] WASC, 2004, Web Application Security Consortium: Threat Classification, Whitepaper. Available from http://www.webappsec.org/projects/threat. Accessed on 03.03.2006

[20] YW Huang, Y Fang, C Hang, CH Tsai, DT Lee, SK Yen, 2004, "Securing web application code by Static Analysis and Runtime protection", Computer Networks 48(2004) pp.788-795

[21] J Joshi, W Aref, A Ghafoor, E Spafford, 2001, Security models for Web based applications, Communications of the ACM 44(2). pp 38-44.

[22] YW Huang, YU Fang, CH Tsai, TP Lin, SK Huang, DT Lee, SY Kuo, 2005, A testing framework for web application security assessment,", Computer Networks 48(2005) pp.739-761.

[23] RR Barton, WJ Hery, P Liu, 2001, An S Vector for web application security assessment. Available from http://209.85.175.104/search?q=cache:3nAQ2kKly0IJ:www.smeal.psu.edu/cdt/ebrcpubs/res_papers/2004_01.pdf+s+vector+%2B+threat+classificaiton&hl=en&ct=clnk&cd=1&gl=in. Accessed on 13.08.2008.

[24] OWASP, 2008, Open Web Application Security Project. Available from www.owasp.org. Accessed on 13.08.2008.

[25] M Peter, R Sasha, S Karen, 2007, A Complete Guide to the Common Vulnerability Scoring System, Version 2.0, Whitepaper. Available from http://www.first.org/cvss/cvss-guide.html. Accessed on 24.07.2008.

[26] CA Alberts,. Dorofee, J Stevens, C Woody,2005, "OCTAVE-S Implementation guide, Volume 1: Introduction to OCTAVE-S, Version 1.0. Whitepaper, Available from http://www.giac.org. Accessed on 23.08.2007.

[27] CENZIC, 2008, Web Application Security: The truth about white box testing Vs. Blackbox testing, Whitepaper. Accessed from http://www.Cenzic.com Accessed on 12.07.2008.

[28] YW Huang, YU Fang, CH Tsai, TP Lin, SK Huang, DT Lee, SY Kuo, 2005, A testing framework for web application security assessment,", Computer Networks 48(2005) pp.739-761.

AUTHORS PROFILE

Dr. K. Ram Mohan Rao hold Post Graduate and Doctoral degree in Computer Science. Presently, he is working as Scientist in Indian Institute of Remote Sensing (NRSC), Dehradun. He has research expertise in the fields of Spatial Databases, GIS Customization and dissemination including programming languages, Location Based Services, Distributed GIS and Risk modelling. His interests include Open source technologies in the field of Geoinformatics. He is a member of Indian Society of Geomatics and Indian Society of Remote Sensing.

Prof. Durgesh Pant hold Post Graduate and Doctoral degree in Computer Science from BIT, Mesra, India. He is now working as Head, Computer Science, Kumaun University, Nainital, India. He has published more than 50 National and International papers in peer reviewed journals, and 3 books of his credit. 11 students have been completed their Ph.D degree under his supervision. He has served as Director, directorate of Counselling & Placement of Kumaun University. He is also the corordinator for Indira Gandhi National Open University, Kumaun University. He is the member of various National and International academic, social and cultural assosociaitons and bodies.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Design and Implementation of GSM based Remote Monitoring and Control system for Industrial process Parameters

Dr.B.Ramamurthy[1]
Associate Professor
Dept of Instrumentation, S K University
Anantapur, Andhra Pradesh India
ramamurthy_ugc@yahoo.com

S.Bhargavi[2]
Assistant Professor
Dept of E&C, SJCIT
Chikballapur, Karnataka, India
bhargavisunil@gmail.com

Dr.R.ShashiKumar[3]
Professor
Dept of E&C, SJCIT
Chikballapur, Karnataka, India
rshashiku@gmail.com

Abstract
The benefits of remote monitoring and control have long been realized in the industrial sector for uses in automation as well as increase of safety/security standards. This led to the emergence of the Remote Monitoring and Control System (RMACS) [1]. In this paper we have developed an integrated Wireless RMACS for monitoring, controlling and accessing the performance of remotely situated device parameters such as Temperature, Pressure, Humidity and Level on real time basis. This paper explores the feasibility of tapping into the short message service component of existing mobile communications network infrastructure, particularly the GSM (Global System for Mobile communication) [2] network to act as a medium for the communication of control signals. From the Wireless RMACS, which is proposed in set up the temperature, pressure, humidity, level could be sufficiently recorded from the remote location, and whenever it crosses the set point the ARM LPC2148 Processor will sends as SMS to a user(s) mobile phone. The user can control the system through his mobile phone. The properly designed Wireless RMACS enable the achievement of truly mobile monitoring, control and other industrial processes; as well as security and safety applications.

Key words:
RMACS, Automation, GSM, ARM LPC2148, Mobile phone, SMS.

1. Introduction

In recent years industrial automation and control systems are an integral part of an industry. Extrapolating automation history forward is an interesting challenge. Today, growth is coming from global expansion and services. A new surge of growth will come through new technology (Wireless) [3], production at the lowest cost for global distribution, and fast time-to-market. A Wireless Industrial Automation communications segment, at the present time, presents a mixture of standardized and proprietary technologies.

Recently, there has been much interest in remote monitoring and control in the field of the Industrial automation. There has also been much interest in wireless communication [4] in industrial sector for uses in automation as well as to increase the safety and security standards. There is a great deal of benefits for process plants to adopt the wireless communication to control systems. Currently the common conditions of use of SCADA systems [5] only allow for control and supervision to take place when the operator and the plant being observed are in the same general vicinity. It led to the emergence of the wireless remote monitoring and control systems. This contribution develops the systematic design methods for the design and implementation of Remote Monitoring and Control system (RMACS) over the wireless communication. This is the main objective and focus of the present research work.

RMACS is on-line real-time monitoring and controlling the field equipment and transmitting the real time picture and test data to the terminal to forecast or diagnose [6]. The RMACS is a effective method to obtain, analyze, transmit, manage and feedback the remote goal information and it combines the most advanced science and technology field of satellite positioning technology, communication technology, Internet technology and other areas, and it is the comprehensive usage of instrumentation, electronic technology, modern communications technology, computer software and so on.

The use of mobile phones or handsets has grown exponentially over the years [7]. As the number of mobile phone users increased, the technology and infrastructure supporting the handsets have also evolved to cope with the traffic created by the number of users. On top of that, the demands of mobile phone users have also changed, and the average mobile phone of today can do many things that would have never been dreamt of 10 years ago.

Voice telephony has always been the primary use of the mobile handset; however, ‘texting’ or text messages were introduced to operate over mobile phone networks and maximize bandwidth utilization. The dominant mobile

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

phone network in the world today is GSM. It is a digital mobile communication network, which developed, rapidly in recent years. This network has coverage in most urban areas and offer support for the SMS [8] that allows users to communicate with each other by sending short text messages to each other at minimal cost. The maximum length of the messages cannot exceed 150 alphanumeric Latin characters. This is enough to send short alert messages or commands to remote system. In the world of automation and control, several methods have thus far been employed to attain remote monitoring and control of various processes. These attempts have met with varying levels of success.

The primary aim of this paper is to propose the concept of a Remote Monitoring and Control System (RMACS) using the combination of a Embedded ARM Processor (ARM7 TDMI-S LPC2148) [9] and a GSM communications module linked by a serial communications port. Using this the process parameters could be efficiently recorded from the remote location and whenever these parameters crosses the set point, the ARM processor will send an SMS alert to a user(s) mobile phone. The user can control the system through the mobile phone enables the achievement of truly mobile monitoring, control and other industrial process; as well as security and safety applications.

The benefits of this paper are:

- Flexibility / modularity in control by the use of an ARM processor.
- Global coverage through the use of the GSM network.
- Extremely low cost device adapted for different applications.
- Scalable, Robust and Reliable.
- Provides password security.
- Efficient and cheap means of communication by use of SMS.
- True mobility using mobile phone sets.

2. Hardware implementation

The Block diagram of the hardware implementation of the entire system is as shown in the Figure 1. All the major subsystem blocks are shown with their inter connections to each module .The block diagram consists of Sensors, ARM TDMI-S LPC2148 Processor, GSM MODEM, Signal conditioning unit, MAX232 Level converter and inverter, Controlling devices, Mobile phone, Line driver ULN2003 and Relay set and Personal computer. In this application, the system was set up to monitor and control the process parameters and ensure that it was within safe operating limits. The detailed descriptions of the components used in the system are explained below.

The system uses an ARM 7 processor to measure and control all operations can be attached to several I/O devices which will in most cases be either of the two:

- Sensory inputs for process control or safety/security monitoring
- Actuators on the outputs to exert control over various aspects of a process.

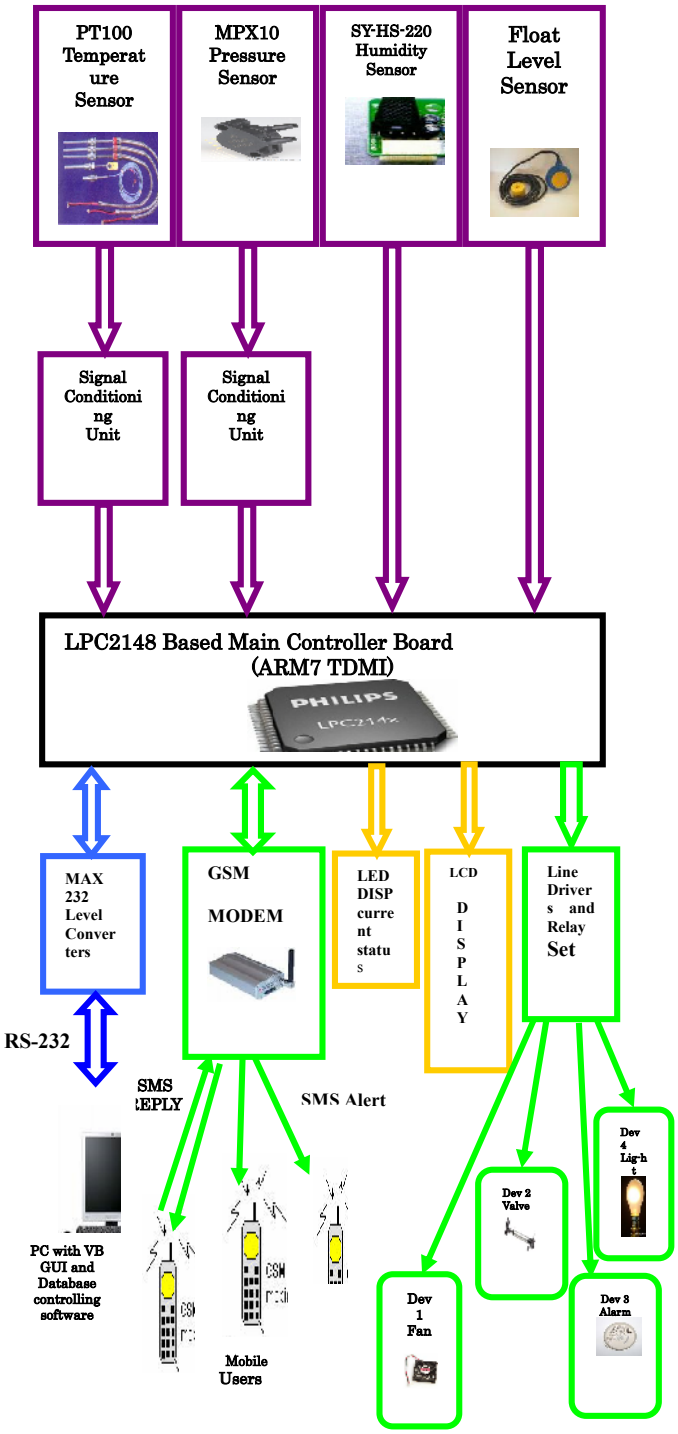


Figure 1: The Block diagram

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

GSM MODEM to provide communication capabilities to the system operated by issuing a string of commands from either the system via the processor or from the user via SMS [11] over mobile handsets. By coupling the two devices, a cost effective, flexible and modular system is achieved. Since, the devices run on relatively low power and the GSM MODEM being a wireless communication module, not only does the system grant mobility to the user, but also the system itself may be operated on a mobile platform in future modifications.

Various Sensor parameters are continuously being monitored by this system. The Sensor output is typically being 'signal conditioned' by means of an Op-Amp etc, which will be able to provide the signals with the right amount of voltages to be properly detected by the ARM processor.

The data logging is achieved continuously by the ARM 7 to the PC via the MAX232 (level converter and inverter). This data is received by the software running on the PC and continuously updates a database.

3. Components or Subsystems Description

3.1 ARM7 TDMI LPC2148 Processor

The NXP (founded by Philips) LPC2148 is an ARM7TDMI-S based high-performance 32-bit RISC Microcontroller with Thumb extensions 512KB on-chip Flash ROM with In-System Programming (ISP) and In-Application Programming (IAP), 32KB RAM, Vectored Interrupt Controller, Two 10bit ADCs with 14 channels, USB 2.0 Full Speed Device Controller, Two UARTs, one with full modem interface. Two I2C serial interfaces, Two SPI serial interfaces Two 32-bit timers, Watchdog Timer, PWM unit, Real Time Clock with optional battery backup, Brown out detect circuit General purpose I/O pins. CPU clock up to 60 MHz, On-chip crystal oscillator and On-chip PLL. Due to their tiny size and low power consumption, LPC2148 are ideal for applications where miniaturization is a key requirement, such as access control and point-of-sale. Serial communications interfaces ranging from a USB 2.0 Full-speed device, multiple UARTs, SPI, SSP to I2C-bus and on-chip SRAM of 8 kB up to 40 kB, make these devices very well suited for communication gateways and protocol converters, soft modems, voice recognition and low end imaging, providing both large buffer size and high processing power. Various 32-bit timers, single or dual 10-bit ADC(s), 10-bit DAC, PWM channels and 45 fast GPIO lines with up to nine edge or level sensitive external interrupt pins make these microcontrollers suitable for industrial control and medical systems.

3.2 Sensors

Sensors are used for process monitoring and for process control. These are essential elements of safe and profitable plant operation that can be achieved only if the proper sensors are selected and installed in the correct locations. In this paper the sensors are used for the measurement of temperature, pressure, humidity and level.

3.2.1 PT100 Temperature Sensor

PT100 [12] is the common abbreviation for the most common type of resistance temperature sensor used in industry. It has a specified resistance of 100.00 ohms at 0°C and is made of Platinum which has an accurately defined resistance vs. temperature characteristic. PT100 sensors were originally made with platinum wire wound on a ceramic former but are now made more cheaply by depositing a platinum film onto a ceramic substrate. PT100 elements are specified over a temperature range of -200°C to 850°C; however the actual operating temperature is determined by the construction of the probe into which they are incorporated. Typical low cost probes are made by soldering the PT100 to PVC or silicon insulated copper wires. Obviously these are limited by the maximum temperature of the insulation.

3.2.2 MPX10 Pressure Sensor

The MPX10 [16] series devices are silicon Piezoresistive pressure sensors providing a very accurate and linear voltage output-directly proportional to the applied pressure. These standard, low cost, uncompensated sensors permit manufacturers to design and add their own external temperature compensation and signal conditioning networks. Compensation techniques are simplified because of the predictability of Motorola's single element strain gauge design.

3.2.3 SY-HS-220 Humidity Sensor

SY-HS-220 [17] Humidity Sensor module consists of a SY-HS-2 sensor and an integrated circuit to provide a linear DC voltage output for 20-95% RH. These modules convert the relative humidity to the output voltage. It operates supply voltage 5V. The features of this are wide humidity operation range, linear DC voltage output, High reliability and long-term stability, easy operation, high sensitivity and low hysteresis. They are specifically designed for use in appliances and controllers.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

3.2.4 Float level Sensor

Level sensors detect the level of substances that flow, including liquids, slurries, granular materials, and powders. All such substances flow to become essentially level in their containers (or other physical boundaries) because of gravity. The substance to be measured can be inside a container or can be in its natural form (e.g. a river or a lake). The level measurement can be either continuous or point values. Continuous level sensors measure level within a specified range and determine the exact amount of substance in a certain place, while point-level sensors only indicate whether the substance is above or below the sensing point. Generally the latter detect levels that are excessively high or low. Float-type sensors [18] can be designed so that a shield protects the float itself from turbulence and wave motion. Float sensors operate well in a wide variety of liquids, including corrosives. When used for organic solvents, however, one will need to verify that these liquids are chemically compatible with the materials used to construct the sensor.

3.3 Signal Conditioning unit

Any instrumentation measurement systems consist of various units starting from sensors to data representation units. Among that signal conditioning is a vital process. This system consists of Amplifiers, Filters, ADC, and DAC etc. The process instrumentation consists of signal conditioning and processing units for very low frequencies. During study of these signals, noise interference is a major problem and complex.

Various Sensor parameters are continuously being monitored by the system. The sensor output is typically being 'signal conditioned' by means of an operational amplifier etc, which will be able to provide the signals with the right amount of voltages to be properly detected by the ARM processor.

The signals from sensors are processed by using ARM7 TDMI LPC2148 processor. The processing unit consists of Amplifier LM324, ADC and a comparator circuit for processing the signals from the sensor. ARM7 TDMI LPC2148 processor includes built in ADC and Comparator.

3.4 GSM MODEM

A GSM modem is a specialized type of modem, which accepts a SIM card, and operates over a subscription to a mobile operator, just like a mobile phone. From the mobile operator perspective, a GSM modem looks just like a mobile phone. A GSM modem can be a dedicated modem device with a serial, USB or Bluetooth connection, or it may be a mobile phone that provides GSM modem capabilities. A GSM modem could also be a standard GSM mobile phone with the appropriate cable and software driver to connect to a serial port or USB port on computer. Any phone that supports the "extended AT command set" for sending/receiving SMS messages, as defined in the ETSI GSM 07.05 Specification can be supported by the Now SMS/MMS Gateway. In the proposed system we have used SIMCOM SIM300 GSM module.

SIM300 [19] is a Tri-band GSM/GPRS engine that works on frequencies EGSM 900 MHz, DCS 1800 MHz and PCS1900 MHz. SIM300 provides GPRS multi-slot class 10 capability and support the GPRS coding schemes CS-1, CS-2, CS-3 and CS-4.With a tiny configuration of 40mm x 33mm x 2.85 mm , SIM300 can fit almost all the space requirement in your application, such as Smart phone, PDA phone and other mobile device. The physical interface to the mobile application is made through a 60 pins board-to-board connector, which provides all hardware interfaces between the module and customers' boards. The SIM300 is designed with power saving technique, the current consumption to as low as 2.5mA in SLEEP mode. The SIM300 is integrated with the TCP/IP protocol, Extended TCP/IP AT commands are developed for customers to use the TCP/IP protocol easily, which is very useful for those data transfer applications.

3.5 Mobile Phone

A mobile phone also known as a wireless phone, cell phone, or cellular telephone is a little portable radio telephone. Mobile Phone can serve as powerful tool for world-wide communication. The Mobile Phone is a natural choice, since it is a communication resource generally available by people, which makes them practically always contactable and capable to send commands to operate the parameters in the industries.

The use of mobile phones or handsets has grown exponentially over the years. As the number of mobile phone users increased, the technology and infrastructure supporting the handsets have also evolved to cope with the traffic created by the number of users. On top of that, the demands of mobile phone users have also changed, and the average mobile phone today can do many things that would have never been dreamt of 10 years ago.

The mobile phone can be used to communicate over long distances without wires. It works by communicating with a nearby base station (sometimes called a "cell") which connects it to the main phone network. As the mobile phone moves around, if the mobile phone gets too far away from the cell it is connected to, that cell sends a message to another cell to tell the new cell to take over the call. This is called a "hand off," and the call continues with the new cell the phone is connected to. The hand-off is done so well and carefully that the user will usually never even know that the call was transferred to another cell. Since a cell phone allows you to be anywhere, and to move around while calling, they became very popular.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

3.6 MAX 232 Level Converters

Since the RS232 is not compatible with today's Microprocessors and Microcontrollers, we need a line driver or voltage converter to convert RS232's signals to TTL voltage levels. One example of such a converter is MAX 232 from Maxim corp. The MAX232 converter converts from RS232 voltage levels to TTL voltage levels and vice versa. One advantage of the MAX232 chip is that it uses a +5v power source, which is the same as the source voltage for the microcontroller. In other words, with a single +5v power supply we can power both the microcontroller and MAX232, with no need for the dual power supplies that are common in many older systems. The MAX 232 has two sets of line drivers for transferring and receiving data.

3.7 Line Driver (ULN 2003)

The ULN2003 [20] is a monolithic high voltage and high current Darlington transistor arrays. It consists of seven NPN Darlington pairs that feature high-voltage outputs with common-cathode clamp diode for switching inductive loads. The collector-current rating of a single darlington pair is 500mA. The Darlington pairs may be paralleled for higher current capability. Applications include relay drivers, hammer drivers, lamp drivers, display drivers (LED gas discharge), line drivers, and logic buffers. The ULN2003 has a 2.7kilo ohms series base resistor for each Darlington pair for operation directly with TTL or 5V CMOS devices. The features are 500mA rated collector current (Single output), High-voltage outputs: 50V, Inputs compatible with various types of logic, Relay driver application.

3.8 Controlling Devices

The devices used to control the industrial process parameters are

- Fan (Temperature)
- Valve (Pressure)
- Alarm (Humidity)
- Bulb (Level)

3.9 Personal Computer

A Personal computer is a programmable machine that receives input, stores and manipulates data, and provides output in a useful format. A personal computer may be a desktop computer, a laptop, a tablet PC, or a handheld PC. The most common microprocessors in personal computers are x86-compatible CPUs. Software applications for personal computers include word processing, spread sheets, data bases, Web browsers and e-mail clients, games, and myriad personal productivity and special-purpose software applications. Modern personal computers often have high-speed or dial-up connections to the Internet allowing access to the World Wide Web and a wide range of other resources. Personal computers may be connected to a local area network (LAN), either by a cable or a wireless connection.

The data logging is achieved continuously by the ARM7 TDMI LPC2148 processor to the personal Computer via the MAX232. This data is received by the software running on the PC and continuously updates a database by using Visual Basic and also we can generate reports and graphs automatically. Focusing on the client requirements, the following capabilities have been provided in the software.

- **Monitoring –** This is the main feature of the system where extracted information is presented for the operator in near real-time. Monitoring has been divided in two sections.
- **Full graphical data representation** – In this section, the user is able to monitor the plant in a very user friendly manner where details are represented in dynamic graphical interfaces in personal computer.
- **Text base data representation** – In this section, the near real time details are represented in tables without graphical objects.

4. Software development

The flowcharts depicting the monitoring and the control process are shown in Figure.2 and Figure.3.

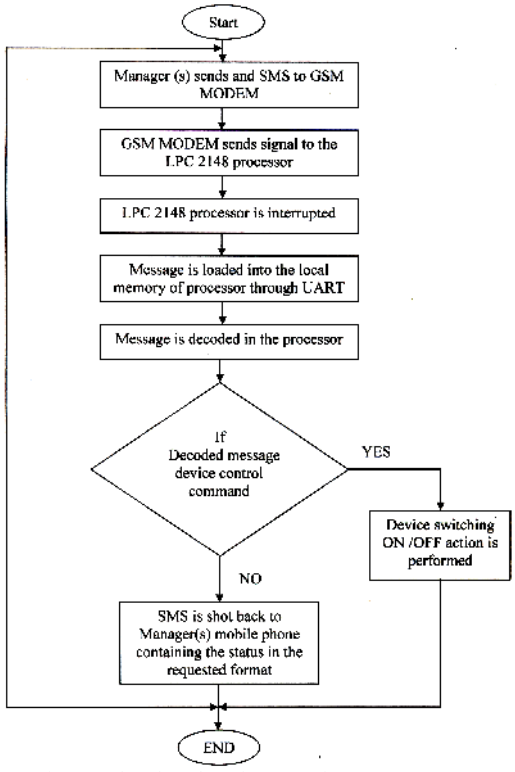


Fig. 2 The Flowchart for control process

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

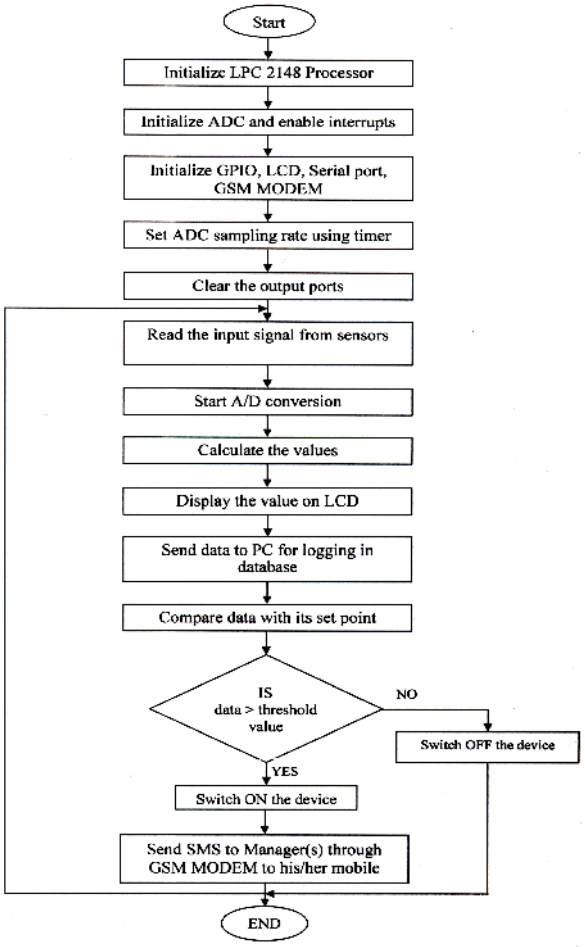


Fig. 3 The flowchart for monitor process

5. Results and discussion

The results obtained by using the proposed system are discussed in this section. Figure 4 shows the measurement and control of Temperature, Pressure, Humidity and Level in the graphical representation. Figure 5 shows the log data in a tabular form. Figure 6 shows the details of current data values and high limits of sensors. Figure 7 shows the status of the parameters in the mobile phone.

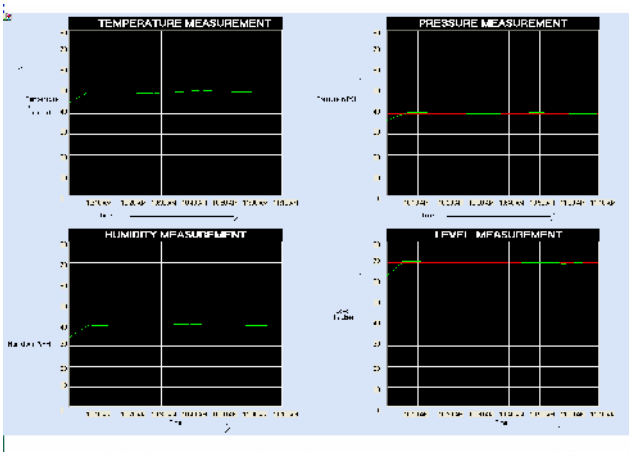


Fig. 4 Graphical Representation of Temperature, Pressure, Humidity and level measurement.

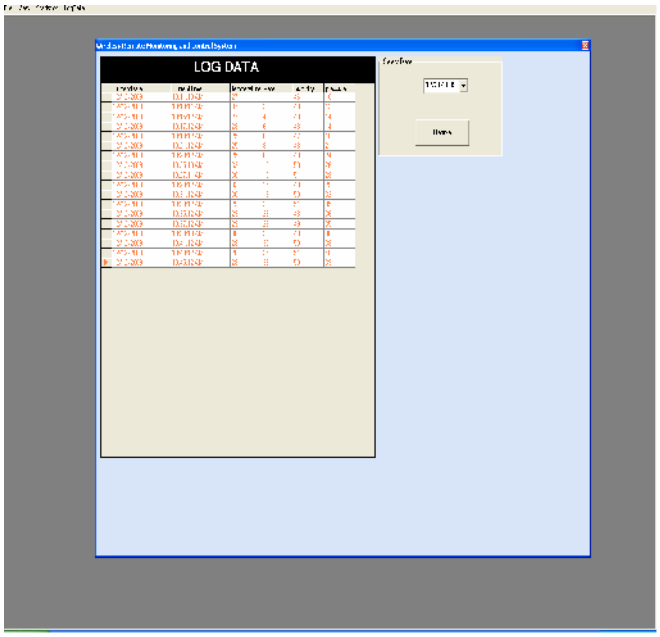


Fig. 5 Log data for industrial process parameters

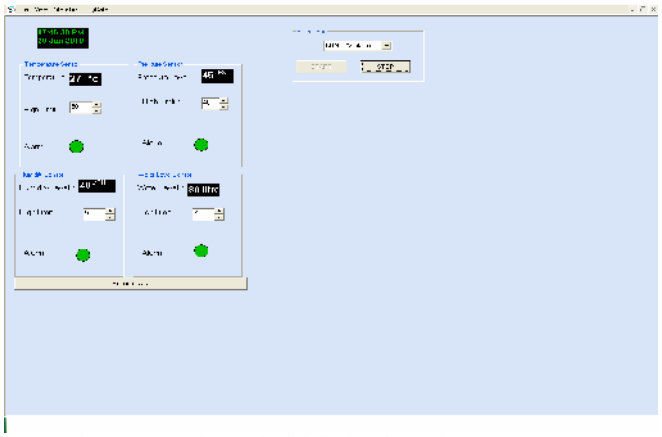


Fig. 6 Current data and High limit values of sensors

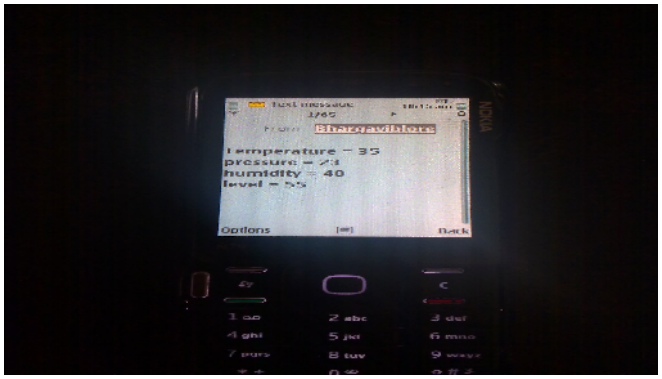


Fig. 7 Status of the Process Parameters in the Remote Mobile Phone User

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

6. Conclusion

The solution has provided a low cost custom built monitoring and controlling system. The use of a ARM Processor, GSM module, Sensors and actuators provide exciting possibilities. However as far as the industrial applications are concerned this can be viewed as a low cost, customized wireless RMACS system. Thus this solution can be customized to suit any other industrial requirement related to monitoring and controlling provided industrial sensors are in use.

In the paper low cost, secure, ubiquitously accessible, auto-configurable, remotely controlled solution for automation of process parameters has been introduced. The approach discussed in the paper is novel and has achieved the target to control process parameters remotely using the SMS-based system satisfying user needs and requirements. GSM technology capable solution has proved to be controlled remotely, provide security and is cost-effective as compared to the previously existing systems. Hence we can conclude that the required goals and objectives of RMACS have been achieved.

References

[1] Chen Peijiang Xuehua, "Design and Implementation of Remote Monitoring System Based on GSM", 2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application
[2] http://www.gsmworld.com
[3] Ciubotaru-Petrescu, B., Chiciudean, D., Cioarga, R., & Stanescu, D,"Wireless Solutions for Telemetry in Civil Equipment and Infrastructure Monitoring", *3rd Romanian-Hungarian Joint Symposium on Applied Computational Intelligence (SACI) May* 25-26, 2006.
[4] Daniel J.S.Lim, Vishy Karri,"Remote monitoring and control of hydrogen Safety Via SMS", 1st International Conference on Sensing Technology November 21-23, 2005 Palmerston North, new Zealand.
[6] N.C Corbett, "Remote monitoring and control of advanced gas turbines", Computing and Control Engineering Journal April 2001.
[7] Ken Wieland, "Spreading the Word" Telecommunications Magazine International Edition October 2004, Issue Highlights.
[8] Puneet Gupta, SMS: How, what and where, Wireless Developer Network
[9] www.phillips.com
[10] Naresh P.Jawarkar, vasif Ahmed, Siddharth A Ladhake, Rajesh D Thakare, "Microcontroller based Remote Monitoring using Mobile through Spoken Commands", Journal of Networks, Vol.3, No.2, February 2008.
[11] Malik Sikandar Hayat Khiyal, Aihab Khan, Erum Shehzadi "SMS Based Wireless Home Appliance Control System (HACS) for Automating Appliances and Security", Issues in Informing Science and Information Technology, Volume 6, 2009.
[12] http:// www.rhopointstruments.com
[13] Malik Sikandar Hayat Khiyal, Aihab Khan, Erum Shehzadi "SMS Based Wireless Home Appliance Control System (HACS) for Automating Appliances and Security", Issues in Informing Science and Information Technology, Volume 6, 2009.
[14] Rifat Shahriyar, Enamul Hoque, S.M.Sohan, Iftekhar Naim, Md.Mostafa Akbar & Masud Karim Khan, " Remote Controlling of Home Appliances using Mobile Telephony", International Journal of Smart Home Vol.2, No.3, July 2008.
[15] http://www.maxim.com
[16] http://www.freescale.com
[17] http://www.hazari.com
[18] http://www.standexelectronics.com
[19] http://www.shoky.com.tw/en/PDF/SIM300.pdf
[20] http://www.st.com

Prof.S.Bhargavi is presently working as a Assistant Professor in the department of Electronics and Communication engg , SJCIT, Chikballapur, Karnataka, India. She is doing PhD in SK University, Anantapur. She is having 12 years of teaching experience. Her areas of interest are Embedded systems, Low Power VLSI, Wireless communication, ASIC and Cryptography.

Dr.B.Rama Murthy is presently working as an Associative Professor in Department of Instrumentation & USIC, SKU, Anantapur, Andhra Pradesh, India. He is having 18 years of Research & teaching experience and 4 years of Industry experience. He guided 5 Ph.D.'s & 4 M.Phil's. His areas of interest are Embedded Systems, Mobile Communication, Smart Cards and Industrial Instrumentation.

Dr. R. Shashikumar is presently working as a Professor in E & C dept, SJCIT, Chikballapur, Karnataka, India. He is having 10 years of teaching and 6 years of Industry experience. His areas of interest includes ASIC, FPGA, Network Security.

SOA based Adaptive Reliable Service Delivery Framework using QoS: A Fuzzy-Bayesian Network Approach

R.Sivaraman
Dy.Director-CCT
Anna University Tiruchirappalli
Email: rsiva.raman@yahoo.com

Dr.RM.Chandrasekeran
Registrar
Anna University Tiruchirappalli
Email: aurmc@sify.com

Abstract
Service Oriented Architecture has gained a universal acceptance as a strategy for developing new applications in dynamic environments through self contained, reusable and configurable services. By adherence to the web services standards stack, these services can interoperate with other services and can also be transformed into a composition of services. The behavior and operation of a service has to be closely monitored for an efficient delivery to the client. Ranking and selection of the services among various service providers have become an important factor for a successful business solution. Quality of Service (QoS) determines the quality and usability of a service and is determined after analyzing the QoS attributes collected from various sources. Hence it has become our research motivation for transfer and reception of QoS data. In this paper, we propose an adaptive and reliable service delivery framework, which adopts a Fuzzy-Bayesian Network (FBN) approach through which the web services can be ranked and selected automatically. Also this framework offers reliability and availability aspects of QoS demand.

Keywords : Web service selection, QoS, Fuzzy-Bayesian Network (FBN), Web Services ranking

1. Introduction

The QoS specification in Web Services has become the need of the day. The service consumers aim to have a good service performance with relatively low waiting time, high availability and reliability to successfully use the service. Also the service providers are formulating their efforts to achieve high throughput guarantees with low response time using dynamic capacity allocation, resource allocation and load balancing in order to cater to a large volume of clients with assured QoS. In the present architecture each of the input requests are manually mapped to a pre-specified category of the provider at the registry. The service consumer has little knowledge about the QoS attributes associated with that particular category and the run-time evaluation of QoS metrics is not available To bridge the gap between the Web Services layer and the underlying QoS aware transport methodologies, a suitable framework may be developed which enables the efficient service selection.

This framework should allow both the clients and service providers to specify their requests and service offers with QoS properties. It should also map the QoS requirements from the Web service and application layer onto the underlying networking layer thereby achieving an overall boost in the QoS across different layers as that of the Internet model. For this a well defined set of QoS indices are used which includes non-runtime data like cost and run-time data as availability, execution time, mean response time, trustworthy, performance, capacity. The cost is a measure of the price involved in requesting the service, which is based on volume of data and execution time is the guaranteed max (or average or min) time taken between the arrival of service request and catering of the request. The capacity represents the limit of concurrent requests for guaranteed performance. For some of the QoS metrics it is not dependable to rely on the service provider's advertisement as it may have a biased version favoring their own service. So the QoS data has to be adjudged in a neutral way and this leads to the development of an independent framework which not only relies on service provider but also considers other indices.

The rest of the paper is organized as follows, the section 2 discusses the survey of literature, section 3 describes the design and architecture of the proposed framework and section 4 deals with the Fuzzy-Bayesian Network. In section 5 we provide the implementation and experimental results. Section 6 discusses the possible future work and it concludes.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

2. Related Work

With the widespread use of Web Services the QoS have became the significant factor in selecting a best service among the providers and it constitutes the differentiating point for a group of services. QoS generally refers to nonfunctional properties of Web services such as price, performance, reliability, integrity, availability, accessibility, interoperability, security, etc. [1]. Certain traditional approaches [2] make use of QoS policies and a weighing system to judge through the attributes. This approach is applicable only when there is a relatively small set of services and attributes. Since all the services cannot be weighed through a common attribute thereby limiting this approach to a smaller set of services. Also rule based model proposed by Patel et.al [3] is too restrictive and provide a condensed set of alternatives for service execution. Menasce [4] in his research has highlighted the need for QoS definition, specification and evaluation in WS from the perspective of both service provider and service user. W3C [5] has summarized the key requirements of QoS for Web services. Recently, several research works have dealt with the definition of QoS languages for WS-based applications. Ludwig [6] in his work have discussed about the QoS languages and the service level agreements(SLA) that are to be initiated between the service consumer and service providers for proper operation of the system after satisfying the QoS requirements. Hewlett-Packard (HP) has proposed a Web Services Management Language (WSML) and framework and IBM have developed a Web Service Level Agreement (WSLA) language.

In a SOA environment the QoS data of the service providers continuously change during the lifetime of a service. There is no mechanism that periodically updates the QoS data into the registries. Al-Masri et al.[7] have stated that about 53% URL's of the Universal Business Registries(UBR) are invalid and they cannot be validated with a valid WSDL document. This means that the metadata information collected at the time of registration of a service in the registry, has been modified and not yet been reported to the registry for proper updation. Since the information contained in the UBR are not accurate, it utilized more resources when performing the binding of web services and wasted a considerable time in trying to communicate with either non functional or poorly functioning web services. The clients should know the revised QoS information to choose the appropriate service accordingly.

Various query based and monitoring based methods exist [8-10] for obtaining the QoS values of the services. The former periodically queries the service and actively request for the QoS information whereas the later utilizes a monitoring engine that runs on the service provider side as a middleware or acts as a central proxy. Lin et.al.[11] have formalized the service selection for a Web service composition as a fuzzy constraint satisfaction problem. Each QoS criteria has five fuzzy sets describing its constraint levels: Poorly Acceptable (PA), Almost Acceptable (AA), Acceptable (A), Very Acceptable (VA), and Extremely Acceptable (EA). The overall QoS preference of a user is then represented by a fuzzy expression which is composed of a group of fuzzy sets connected by the and logical operator. They have introduced necessary definitions for mapping a Web service dynamic composition to a fuzzy constraint satisfaction problem (FCSP) and finding its solution through NP-hard [12]. Xiong et.al. [13] modeled QoS based service selection as a fuzzy multiple criteria decision making problem (FMCDM). QoS criteria are evaluated by a set of linguistic expressions L1 = {Very Poor (VP), Medium Poor (MP), Poor (P), Medium (M), Good (G), Medium Good (MG), Very Good (VG)}. Similarly, the weights of QoS criteria are expressed by a set of linguistic expression L2 = {Very Low (VL), Low (L), Medium (M), High (H), Very High (VH)}. These linguistic expressions are mapped to fuzzy set membership functions to yield the results. Agarwal et.al [14] suggested that user preferences can be defined as a set of fuzzy If-Then rules. The If part contains membership functions of various QoS criteria of a Web service and the Then part is one of the membership functions of a special concept representing the rank of the Web service. In each rule, fuzzy concepts like fast, medium, slow or cheap, medium, expensive are used to express imprecise values of QoS properties and these fuzzy concepts are modeled as membership functions.

3. Proposed Framework

In this paper we propose to create a reliable architectural framework that can manage faults and can adapt itself to provide a reliable service. This framework decouples the clients from the service providers and takes care of the service execution after due consideration of all QOS issues. Presently the web service clients are typically hardwired with the service providers and does not adapt dynamically in case of service failures. Unavailability of a service is handled only with a manual intervention during failures and fixing them accordingly. This involves service stoppage which directly hinders the performance of the system. The proposed framework addresses this issue by acting as a layer in between the web service clients and the web service providers and dynamically reconfigures itself to respond to the faults so that the business processes can continue in the event of service failures. By virtue of using SOA and XML based web services, this framework is multiplatform enabled along with multi protocol support that can bridge various interfaces, thereby removing the difficulty of integrating mobile and wireless sensor network clients with this framework. The clients submit their requests to the framework and during runtime the

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

framework will select on behalf of the clients the most appropriate service available or its equivalent compatible service depending on the QOS parameters. As shown in Fig. 1, the modules of the proposed framework include a Service Controller (SC),

regular Internet link. In case of failure of any of the link the other takes control over thereby offering the client a reliable connection path. The route redirector acts as a cache and redirection router in case of any link failure to re-route the reply through any of the available clients. Since all the clients are interlinked together, the reply for Client-2 can be redirected and pushed through Client-4 in case of failure of connectivity between Client-2 and the server.

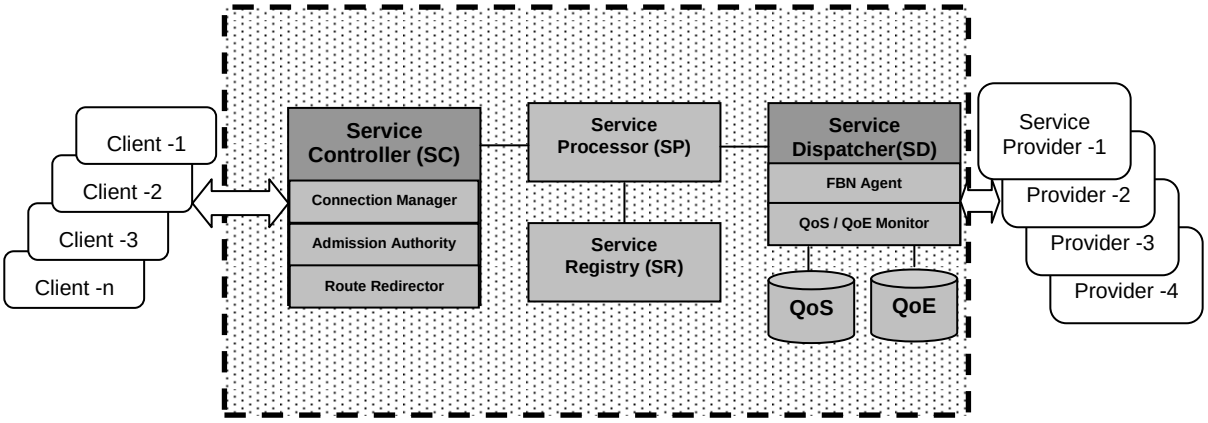


Fig.1. Proposed Framework

Service Processor (SP), Service Registry (SR) and a Service dispatcher (SD) which holds a QoS and QoE repository. All the requests from the clients are routed through this framework to the services located in the servers. The clients search for the available services through this framework, get the WSDL and start communicating with the desired service. The role of SP is for registering and configuring all the available and compatible services from various service providers. Contractual relationships are established with the service providers and after due testing, the services are registered and configured into this framework. All the available services are stored in SR that acts as a private UDDI registry. The SR also holds additional information like the number of retries in case of a service access failure and the list of equivalent services available with other service providers.

The SC acts as an admission authority that blocks the unauthorized SOAP messages from the clients. The clients authenticate themselves by their physical hardware address which has been previously registered with the SC. The list of all registered clients and their respective hardware addresses are stored in the SC which are validated during service requests and only authorized clients are permitted to invoke service requests. The SC also holds a connection manager which holds a multiple connection links with the client routed through different Internet Service Providers (ISPs). In our experimental study each client is connected with three different links, a Virtual Private Network(VPN) link, a Multi Protocol Labeled Switching (MPLS) link and a

The service dispatcher (SD) is responsible for the transfer of request from the service consumer to the service provider. It makes use of the Fuzzy-Bayesian Network algorithm for efficient judging of the QoS and QoE values. These values are collected from the repositories and continuously monitored for the dynamic changes. The services are ranked accordingly and the service requests from the clients will be directed to the available provider for service invocation.

4. Fuzzy-Bayesian Network Approach

Fuzzy sets were derived from the concepts proposed by L.A. Zadeh. Bayesian network is a model of uncertain knowledge representation and reasoning based on the probability and graph theory and was proposed by J. Pearl [15].Fuzzy Bayesian networks (FBN) represent a machine learning model comprising of variables which are simultaneously fuzzy and uncertain. A FBN is a Bayesian network which has fuzzy variables [16]. In the proposed work, Y represents the services and X represents the QoS metrics. By the Bayesian equation [17], if X and Y are events, and $P(Y)>0$, then the conditional probability of X , given Y is,

$$P(X \mid Y) = \frac{P(XY)}{P(Y)}$$

$$P(XY) = P(Y)P(X \mid Y) = P(X)P(Y \mid X)$$

Assume that $\{Y_i, i \in n\}$is a countable unit of events. Let Y be another event and suppose that we know $P(Y_i)$ and

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

$P(X \mid Y_i)$ for each $i \in n$. Accordingly the total probability formula is,

$$P(Y)=\sum_{i=1}^{n} P(Y_i)P(X \mid Y_i)$$

Bayesian equation is

$$P(Y_i \mid X_j) = \frac{P(X_j Y_i)}{P(X_j)}$$

$$P(Y_i \mid X_j) = \frac{P(Y_i)P(X_j \mid Y_i)}{\sum_{i-1}^{n} P(Y_i)P(X_j \mid Y_i)}$$

for i=1 to n, j=1 to m

Assuming that E_1, E_2, ….. E_n are events in the experiment, the multiplication rule of probability is given by

$$P(E_1)P(E_2 \mid E_1).....P(E_N \mid E_1, E_2, ...E_{N-1}) .$$

If the prior probability $P(Y_i)$ and the prior conditional probability $P(X_j \mid Y_i)$ is known then, $P(Y_i \mid X_j)$can be derived from the Bayesian Equation. According to the derivations of the fuzzy subset [18,19,20,21], any function $\mu_{\tilde{A}}$ which maps the domain U to the closed interval [0,1], ($\mu_{\tilde{A}}$:U→[0,1]) ascertains a fuzzy subset Ã. The function $\mu_{\tilde{A}}$ is a membership function of the fuzzy subset and so $\mu_{\tilde{A}}(u)$ is the membership grade of u for Ã and its scalar probability is:

$$P(\tilde{A}) = \sum x \in x \mu_{\tilde{A}}(x) P(x) .$$

Using the above equation the fuzzy Bayesian equation can be expressed as,

$$P(\tilde{Y} \mid X) = \sum_{i \in I} \sum_{j \in J} \mu_{\tilde{Y}}(Y_i) \mu_{\tilde{X}}(X_j) P(X_j \mid Y_i) P(X_i) \Big/ P(\tilde{X})$$

5. Implementation and Experimental Results

The proposed framework was implemented in C# and performed on 4 servers acting as service providers. The servers had an Intel Quad Core processor with 8GB RAM running on a Windows2008 operating system. One of them acted as the Service Dispatcher that monitors and hosts the QoS and QoE repositories. 16 clients were connected to this framework each located over a distance of 200Kms from the servers. The clients were composed of Intel Core2Duo processor with 2GB RAM having Windows XP operating system. The clients and servers were connected on a MPLS-VPN framework with a backup VPN connectivity and a redundant Internet line from different ISPs. This experimental setup is created to simulate an e-business environment with the clients and offices spaced in different locations. The QoS metrics of cost, availability, execution time, mean response time, trustworthy, performance and capacity are measured and they represent the inter-organizational transactions only. Suitable Fuzzy Bayesian rules were created and the model was tested in the Matlab was shown in Fig.2

6. Conclusion

The quality and eventually the performance of a SOA system depends upon the selection of a suitable service from the available services after giving due consideration on the QoS requirements. Since the components of a typical SOA application are built by integration from various service providers in a distributed environment, the QoS requirements change in runtime. In fact the quality

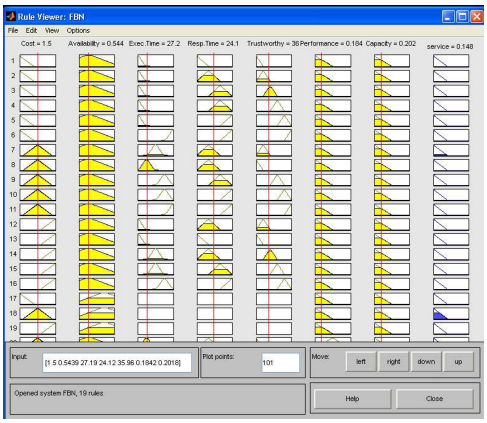


Fig. 2. FBN Rules

also relies on the underlying support systems and on the network resources. In this paper we analyzed the Web services selection based on the Fuzzy-Bayesian Network model and the experimental results showed a considerable increase in performance and reliability. This work may open up the doors towards a broader analysis which may include many other approaches along with FBN model to improve the selection of Web services.

References

1. Understanding Quality of Services for Web Services and Web Services OoS Requirements, http://www.ibm.com/developerworks/library/wsquality.html
2. Using Policy-Based QoS to Enable and Manage WMM in Enterprise Wireless Deployments, http://blogs.msdn.com/wndp/archive/2006/06/30/653047.aspx
3. C. Patel, K. Supekar, Y. Lee, 2003, A QoS Oriented Framework for Adaptive Management of Web Service based Workflows, Lecture Notes in Computer Science (LNCS), Volume 2736, Springer-Verlag, ISBN 3-540-40806-1, pp. 826 – 835.
4. D. A. Menasce. QoS Issues in Web Services. *IEEE Internet Computing*, 6(6):72–75,Nov./Dec.2002.
5. Qos for web services: Requirements and possible approaches.W3C Working Group Note, 25 November 2003, Available at http://www.w3c.or.kr/kr-office/TR/2003/ws-qos/.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

6. H. Ludwig. Web services qos: External SLAs and internal policies or: How do we deliver what we
7. promise? In Proceedings of the Fourth International Conference on Web Information Systems Engineering Workshops (WISE'03). IEEE, 2004.
8. E. Al-Masri and Q. H. Mahmoud, "Investigating Web Services on the World Wide Web," Int'l World Wide Web Conf., April, 2008,Beijing, China,pp.795-803.
9. L. Zeng, H. Lei, and H. Chang, "Monitoring the QoS for Web Services" Proc. Int'l Conference on Service-Oriented Computing, LNCS 4749, 2007, pp. 132–144.
10. Z. Du and J. Huai, "Research and Implementation of an Active Distributed Web Service Registry," Journal of Software, 2006,17(3),454-462
11. F. Li, F. Yang, et al., "A Policy-driven Distributed Framework for Monitoring Quality of Web Services," Proc. IEEE Int'l Conf. on Web Services, 2008, pp.708-715
12. M. Lin, J. Xie, H. Guo, and H. Wang. Solving Qos-Driven Web Service Dynamic Composition as Fuzzy Constraint Satisfaction. In Proc. of the IEEE International Conference on E-Technology, E-Commerce and EService, IEEE Computer Society, pp. 9–14, 2005.
13. D. Dubois, H. Fargier, and H. Prade. Possibility Theory in Constraint Satisfaction Problems: Handling Priority, Preference and Uncertainty. In Journal of Applied Intelligence, Vol. 6(4), pp. 287–309, 1996.
14. P. Xiong and Y. Fan. QoS-Aware Web ServiceSelection by a Synthetic Weight. In Proc. of the 4th International Conference on Fuzzy Systems and Knowledge Discovery, pp.632–637, 2007.
15. S. Agarwal and S. Lamparter. User Preference Based Automated Selection of Web Service Compositions. In Proc. of the ICSOC Workshop on Dynamic Web Processes, Amsterdam, Netherlands, 2005.
16. J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann, San Mateo, CA, 1988
17. Christopher Fogelberg, Vasile Palade, and Phil Assheton. Belief propagation in fuzzy Bayesian networks. In Ioannis Hatzilygeroudis, editor, 1st International Workshop on Combinations of Intelligent Methods and Applications(CIMA) at ECAI'08, pages 19–24, University of Patras, Greece, 22 July 2008.
18. Guangzu Liu, Probability Theory and Applied Mathematical, Higher Education Press, Beijing, 2000
19. Giulianella Coletti and Romano Scozzafava, "Conditional probability, fuzzy sets, and possibility: a unifying view",Fuzzy Sets and Systems, 2004, volume 144, issue 1, pp.227-249.
20. Xingui He, Fuzzy Theories and Fuzzy Techniques in Knowledge Processing, National Defence Industry Press,Beijing, 1998.
21. Shiyong Li, Engineering Fuzzy Mathematics with Applications, Harbin Institute of Technology Press, Harbin, 2004.
22. Hao Tang, Shi Liu, Basic Theory of Fuzzy Bayesian Networks and Its Application in Machinery Fault Diagnosis, Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007).

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Feasibility Analysis of using Cannon Method to Perform Block Rhotrix Multiplication.

Abdullahi M.[1], Ezugwu A. E.[2], Ahmed B. I.[3], Tukur S. M.[4]

Department of Mathematics,

Ahmadu Bello University,

Zaria, Nigeria.
1 muham08@yahoo.com, 2 code_abs@yahoo.com, 3 barroonia@yahoo.co.uk, 4 smtexpert@yahoo.com

Abstract-This paper introduces the concept of parallel rhotrix – rhotrix multiplication, algorithm design and analysis based on the idea of parallelization of matrix–matrix multiplication with Cannon algorithm.

Keywords: Rhotrix, rhotrix blocking, alignment operation, shift operation, main entries, heart entries.

1. INTRODUCTION

Rhotrices multiplication was first introduced and defined by Ajibade[1], and later Sani [2] gave an alternative row-column multiplication of higher dimensional rhotrices. Their work was basically an idea that offers exercises for mathematical enrichment with sets of objects which are in some ways between 2-dimensional vectors and 2x2-dimensional matrices as in [1], and for those objects which are in some ways between nxn-dimensional matrices and (2n-1) x (2n-1) dimensional matrices as in [3].

In this paper our interest is based on n-dimensional rhotrix operation and its multiplication. The name rhotrix come up as a result of the rhomboid nature of the arrangement of some mathematical arrays, which can simply be seen as the combination of matrices. A rhotrix R is called a real rhotrix if all its entries belong to the set of Real Numbers. On the other hand a rhotrix R is called an integer rhotrix if all its entries belong to the set of Integer Numbers.

Rhotrix is a matrix like array of n- dimensional form of rhomboid nature. The major difference between matrix and Rhotrix is the arrangement and the introduction of concept of a heart (h).

$$\left\langle \begin{array}{ccc} & a & \\ b & h(R) & d \\ & e & \end{array} \right\rangle$$

Fig 1: Rhotrix Representation

Rhotrix can be split into two forms of matrices: the main entries and the heart as shown in the figure below.

$$\begin{bmatrix} a & d \\ b & e \end{bmatrix} \qquad [h(R)]$$

where h(R) is the heart of the rhotrix.

Fig 2: Rhotrix main entries and the heart

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

2. HIGHER-DIMENSIONAL RHOTRICES

Ajibade[1] further added that, the dimension of rhotrices can be expanded in size but would always be of odd dimension, he further proved that an n-dimensional rhotrix Rn will have |Rn| entries where |Rn| = (1/2)(n2+1). n-dimensional rhotrix can be represented as shown in figure 3.

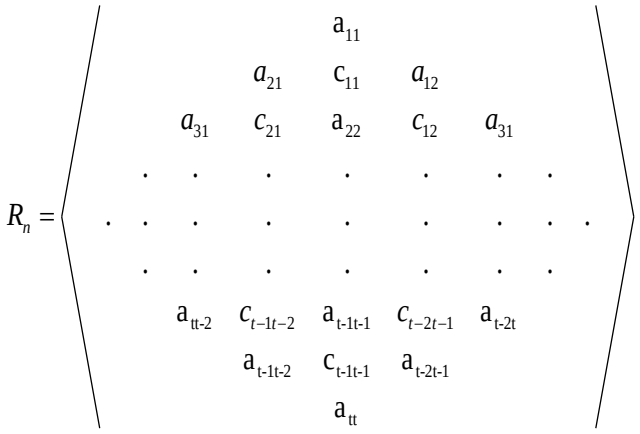


Figure 3: General representation of n-dimensional rhotrix (adopted from [3])

Where t=(n+1)/2, so that if n=3, then t=2. If n=5, t=3 and if n=7, t=4. In this case the 3rd, 5th and 7th dimensional rhotrices are given in figure4:

$$R_3 = \begin{pmatrix} & a_{11} & \\ a_{21} & c_{11} & a_{12} \\ & a_{22} & \end{pmatrix}$$

$$R_5 = \begin{pmatrix} & & a_{11} & & \\ & a_{12} & c_{11} & a_{12} & \\ a_{31} & c_{12} & a_{22} & c_{12} & a_{13} \\ & a_{31} & c_{22} & a_{23} & \\ & & a_{33} & & \end{pmatrix}$$

$$R_7 = \begin{pmatrix} & & & a_{11} & & & \\ & & a_{21} & c_{11} & a_{12} & & \\ & a_{31} & c_{21} & a_{22} & c_{12} & a_{13} & \\ a_{41} & c_{31} & a_{32} & c_{22} & a_{23} & c_{13} & a_{14} \\ & a_{42} & c_{32} & a_{33} & c_{23} & a_{24} & \\ & & a_{43} & c_{33} & a_{34} & & \\ & & & a_{44} & & & \end{pmatrix}$$

We define the main rows and columns of rhotrix R_n depicted in figure 3 as:

$$\begin{matrix} a_{t,1} & a_{t-1,1} & . & . & . & a_{1,1} & & & \\ & a_{t,2} & a_{t-1,2} & . & . & . & a_{1,2} & & \\ & & a_{t,3} & a_{t-1,3} & . & . & . & a_{1,3} & \\ & & & . & . & . & . & . & . \\ & & & & . & . & . & . & . & . \\ & & & & & a_{t,t} & a_{t-1,t} & . & . & . & a_{1,t} \end{matrix}$$

and

$$\begin{matrix} & & & a_{1,1} & . & . & . & a_{1,t-1} & a_{1,t} \\ & & a_{2,1} & . & . & . & a_{2,t-1} & a_{2,t} & \\ & a_{3,1} & . & . & . & a_{3,t-1} & a_{3,t} & & \\ . & . & . & . & . & . & & & \\ . & . & . & . & . & . & & & \\ a_{t,1} & . & . & . & a_{t,t-1} & a_{t,t} & & & \end{matrix}$$

Respectively, the rows and columns of R_n, i.e. those containing the values c_{lk} are similarly defined likewise.

3. MULTIPLICATION OPERATION OF RHOTRICES

Multiplication operation denoted by "o" can be defined in many ways, but in this paper we restrict our idea based on B. Sani's [3] concept of row-column multiplication of high dimensional rhotrices. Representing the n-dimensional rhotrix in figure 3 above by $R_n = (a_{ij},c_{lk})$ where a_{ij} and c_{lk} represent the a_{ij} and c_{lk} elements respectively, with i,j=1,2,3,4,……,t and l,k=1,2,3,4,…..t-1.

The general idea of multiplying any two rhotrices of the same dimension still remain the same with that which is obtainable in the matrix row-column multiplication method where the set of elements a_{ij} multiply each other and elements c_{lk} also do likewise. Thus, the multiplication of any two rhotrices say R_n and Q_n can defined as

$$R_n \circ Q_n = (a_{i1j1,}c_{l1k1}) \circ (b_{i2j2,}d_{l2k2}) =$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

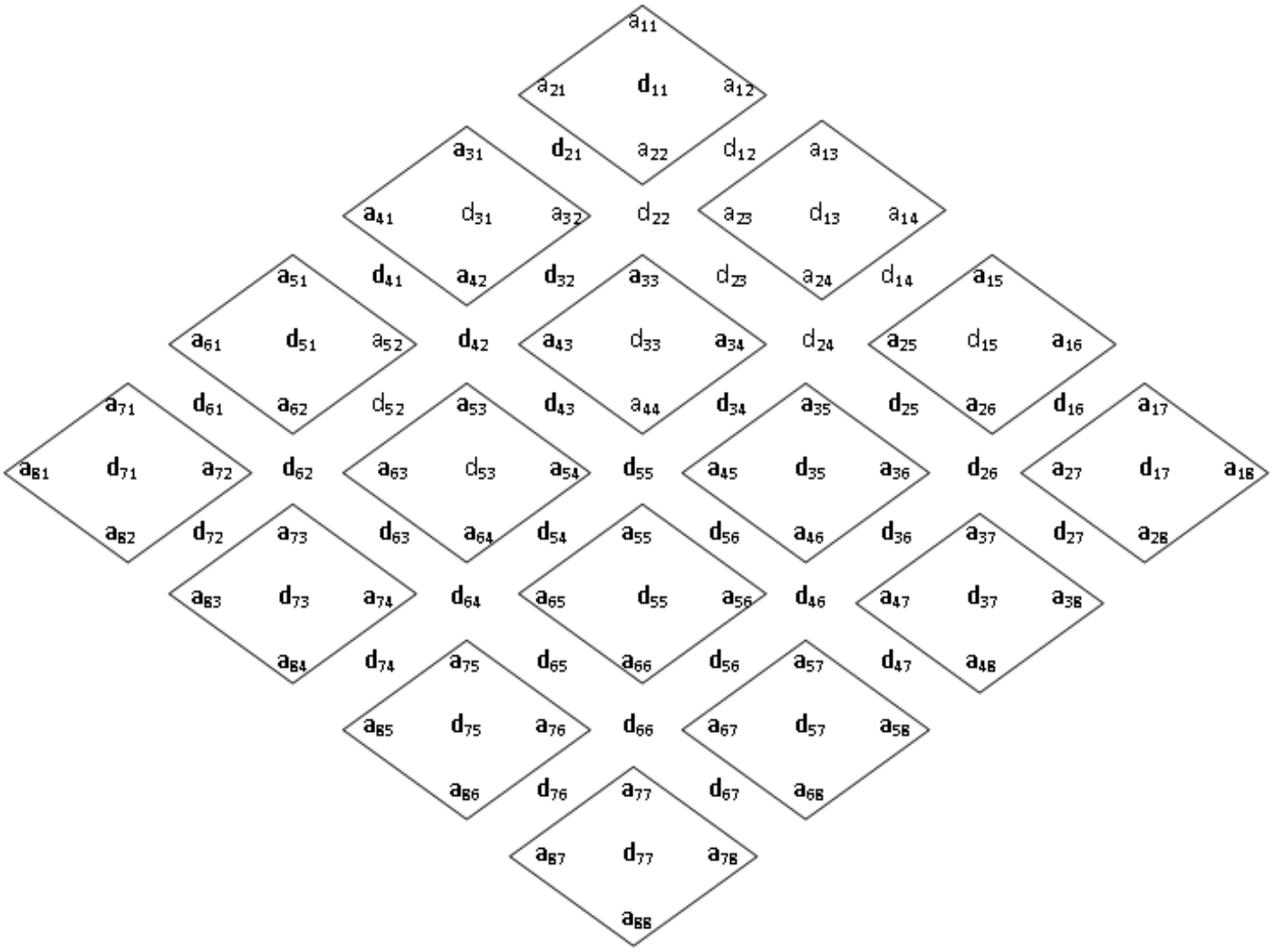


Figure 6a: Blocking process of an eight dimensional rhotrix with main entries **a** and hearts **d**

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

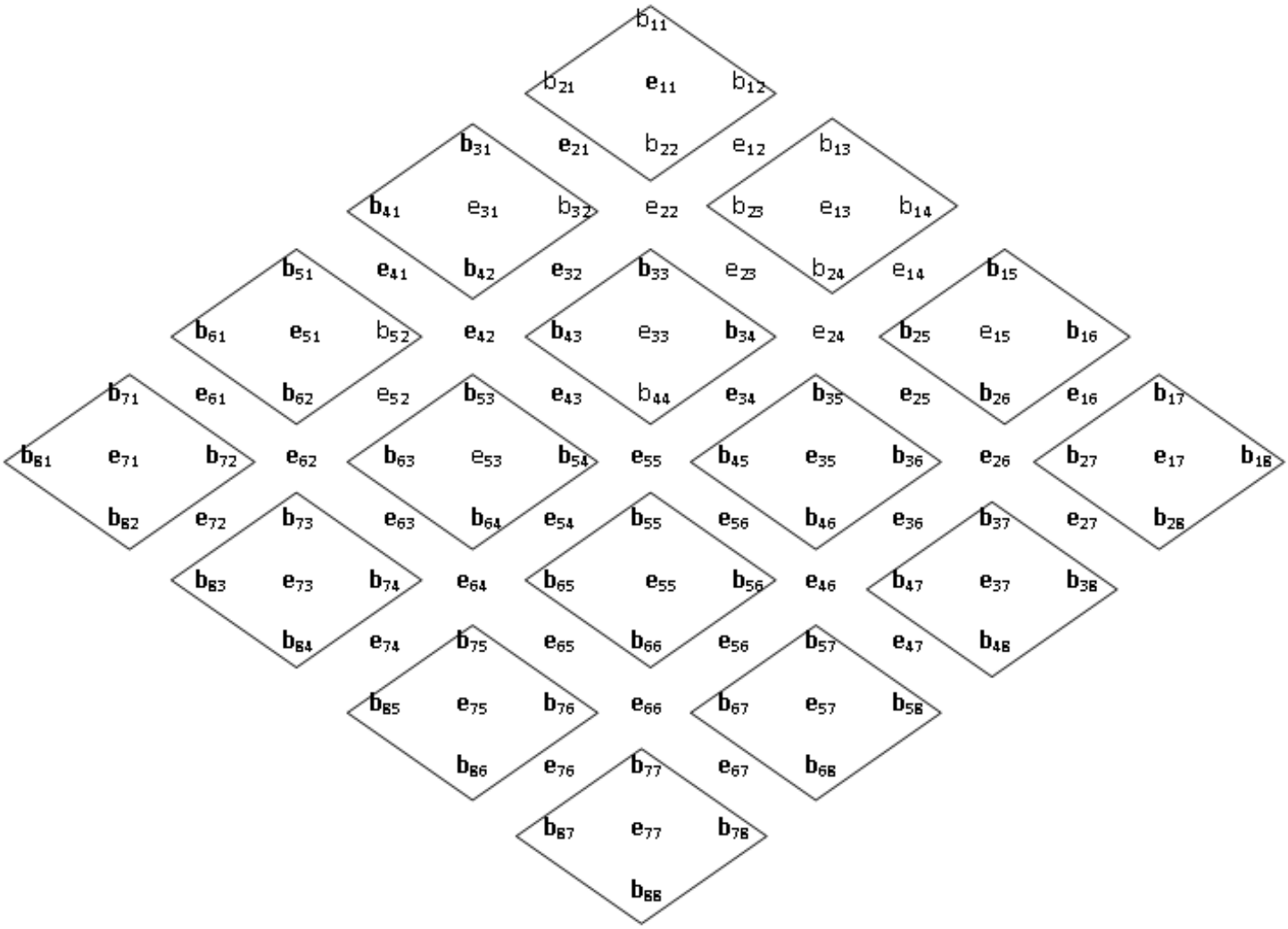


Figure 6b: Blocking process of an eight dimensional rhotrix with main entries **b** and hearts **e**

The blocking process from figure 6a and 6b above was obtained in similar manner as that of matrix blocking method performed using cannon algorithm. The eight dimensional rhotrix above was blocked into sixteen partitions comprising of rhotrix main entries and the heart elements denoted by **a, b** and **d, e**. Those elements outside the blocks are mapped to their corresponding elements with the same index number thereby allowing them to belong to the same partitioned block.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Two rhotrices R (A) and R (B) are considered to illustrate multiplication as shown below:

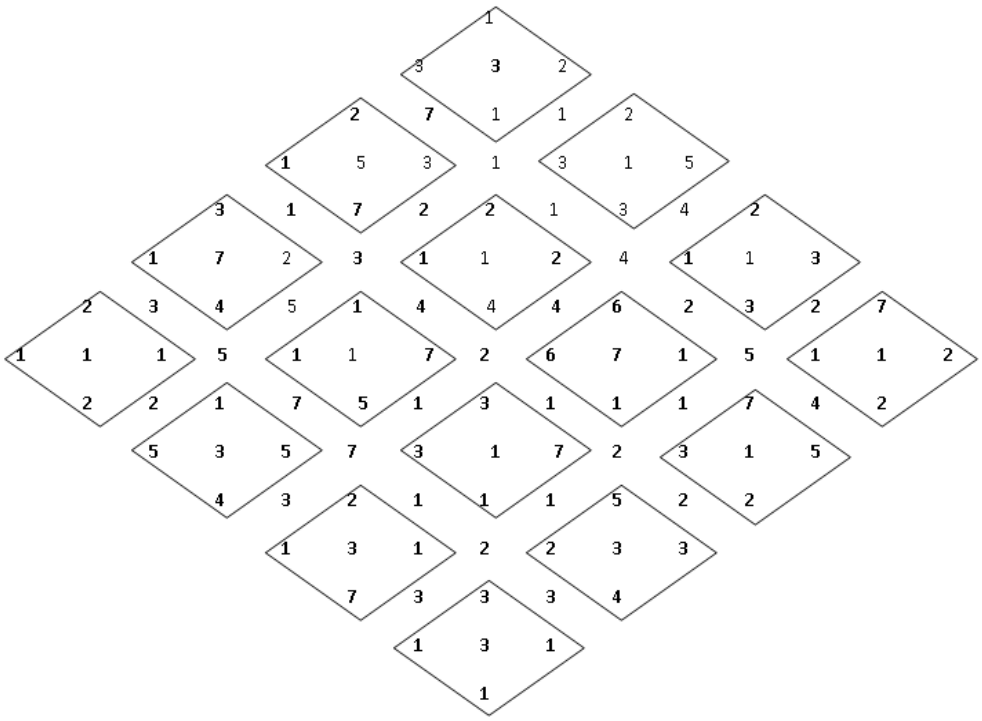


Figure 7a: Example of blocking process of an eight dimensional rhotrix, R (A)

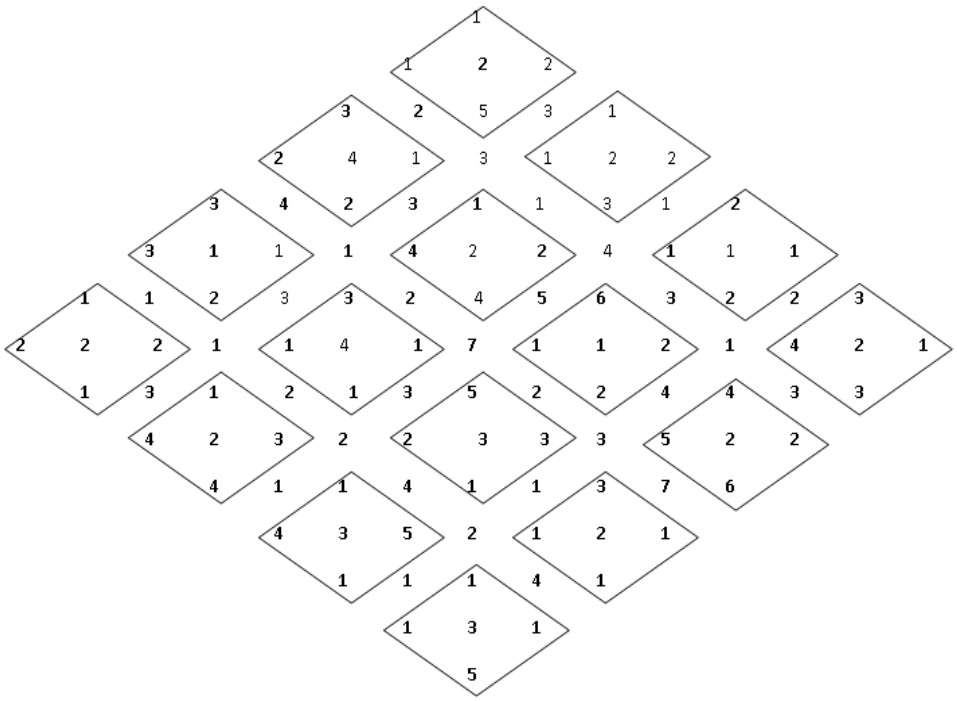


Figure 7b: Example of blocking process of an eight dimensional rhotrix R (B)

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

$$A=\begin{bmatrix}\begin{bmatrix}1&2\\3&1\end{bmatrix}&\begin{bmatrix}2&5\\3&3\end{bmatrix}&\begin{bmatrix}2&3\\1&3\end{bmatrix}&\begin{bmatrix}7&2\\1&2\end{bmatrix}\\\begin{bmatrix}2&3\\1&7\end{bmatrix}&\begin{bmatrix}2&2\\1&4\end{bmatrix}&\begin{bmatrix}6&1\\6&1\end{bmatrix}&\begin{bmatrix}7&5\\3&2\end{bmatrix}\\\begin{bmatrix}3&2\\1&4\end{bmatrix}&\begin{bmatrix}1&7\\1&5\end{bmatrix}&\begin{bmatrix}3&7\\3&1\end{bmatrix}&\begin{bmatrix}5&3\\2&4\end{bmatrix}\\\begin{bmatrix}2&1\\1&2\end{bmatrix}&\begin{bmatrix}1&5\\5&4\end{bmatrix}&\begin{bmatrix}2&1\\1&7\end{bmatrix}&\begin{bmatrix}3&1\\1&1\end{bmatrix}\end{bmatrix}$$

Figure 8a: Block of matrices of main entries of rhotrix R (A)

$$D=\begin{bmatrix}\begin{bmatrix}3&1\\7&1\end{bmatrix}&\begin{bmatrix}1&4\\1&4\end{bmatrix}&\begin{bmatrix}1&2\\2&5\end{bmatrix}&\begin{bmatrix}1\\4\end{bmatrix}\\\begin{bmatrix}5&2\\1&3\end{bmatrix}&\begin{bmatrix}1&4\\4&1\end{bmatrix}&\begin{bmatrix}7&1\\1&2\end{bmatrix}&\begin{bmatrix}1\\2\end{bmatrix}\\\begin{bmatrix}7&5\\3&5\end{bmatrix}&\begin{bmatrix}1&5\\7&7\end{bmatrix}&\begin{bmatrix}1&1\\1&2\end{bmatrix}&\begin{bmatrix}3\\3\end{bmatrix}\\\begin{bmatrix}1&2\end{bmatrix}&\begin{bmatrix}2&3\end{bmatrix}&\begin{bmatrix}3&3\end{bmatrix}&\begin{bmatrix}3\end{bmatrix}\end{bmatrix}$$

Figure 8b: Block of matrices of heart entries of rhotrix R (A)

$$B=\begin{bmatrix}\begin{bmatrix}1&2\\1&5\end{bmatrix}&\begin{bmatrix}1&2\\1&3\end{bmatrix}&\begin{bmatrix}2&1\\1&2\end{bmatrix}&\begin{bmatrix}3&1\\4&3\end{bmatrix}\\\begin{bmatrix}3&1\\2&2\end{bmatrix}&\begin{bmatrix}1&2\\4&4\end{bmatrix}&\begin{bmatrix}6&2\\1&1\end{bmatrix}&\begin{bmatrix}4&2\\5&6\end{bmatrix}\\\begin{bmatrix}3&1\\3&2\end{bmatrix}&\begin{bmatrix}3&1\\1&1\end{bmatrix}&\begin{bmatrix}5&3\\2&1\end{bmatrix}&\begin{bmatrix}3&1\\1&1\end{bmatrix}\\\begin{bmatrix}1&2\\2&1\end{bmatrix}&\begin{bmatrix}1&3\\4&4\end{bmatrix}&\begin{bmatrix}1&5\\4&1\end{bmatrix}&\begin{bmatrix}1&1\\1&5\end{bmatrix}\end{bmatrix}$$

Figure 8c: Block of matrices of main entries of rhotrix R (B)

$$E=\begin{bmatrix}\begin{bmatrix}2&3\\2&3\end{bmatrix}&\begin{bmatrix}2&1\\1&4\end{bmatrix}&\begin{bmatrix}1&2\\3&1\end{bmatrix}&\begin{bmatrix}2\\3\end{bmatrix}\\\begin{bmatrix}4&3\\4&1\end{bmatrix}&\begin{bmatrix}2&5\\2&7\end{bmatrix}&\begin{bmatrix}1&4\\2&3\end{bmatrix}&\begin{bmatrix}2\\7\end{bmatrix}\\\begin{bmatrix}1&3\\1&1\end{bmatrix}&\begin{bmatrix}4&3\\2&2\end{bmatrix}&\begin{bmatrix}2&1\\4&2\end{bmatrix}&\begin{bmatrix}2\\4\end{bmatrix}\\\begin{bmatrix}2&3\end{bmatrix}&\begin{bmatrix}2&1\end{bmatrix}&\begin{bmatrix}3&1\end{bmatrix}&\begin{bmatrix}3\end{bmatrix}\end{bmatrix}$$

Figure 8d: Block of matrices of heart entries of rhotrix R (B)

Computation at Process P_{00}, C_{00} and H_{00}

After the first alignment

$$C_{00}=A_{00}+B_{00}=\begin{bmatrix}1&2\\3&1\end{bmatrix}\times\begin{bmatrix}1&2\\1&5\end{bmatrix}=\begin{bmatrix}3&12\\4&11\end{bmatrix}$$

$$H_{00}=D_{00}+E_{00}=\begin{bmatrix}3&1\\7&1\end{bmatrix}\times\begin{bmatrix}2&3\\2&3\end{bmatrix}=\begin{bmatrix}8&12\\16&24\end{bmatrix}$$

After the first shift

$$C_{00}=C_{00}+A_{01}\times B_{10}=\begin{bmatrix}3&12\\4&11\end{bmatrix}+\begin{bmatrix}2&5\\3&3\end{bmatrix}\times\begin{bmatrix}3&1\\2&2\end{bmatrix}$$

$$=\begin{bmatrix}3&12\\4&11\end{bmatrix}+\begin{bmatrix}16&12\\15&9\end{bmatrix}=\begin{bmatrix}19&24\\19&20\end{bmatrix}$$

$$H_{00}=H_{00}+D_{00}\times E_{00}=\begin{bmatrix}8&12\\16&24\end{bmatrix}+\begin{bmatrix}1&4\\1&4\end{bmatrix}\times\begin{bmatrix}4&3\\4&1\end{bmatrix}$$

$$=\begin{bmatrix}8&12\\16&24\end{bmatrix}+\begin{bmatrix}20&7\\20&7\end{bmatrix}=\begin{bmatrix}28&19\\36&31\end{bmatrix}$$

After the second shift

$$C_{00}=C_{00}+A_{02}\times B_{20}=\begin{bmatrix}19&24\\19&20\end{bmatrix}+\begin{bmatrix}2&3\\1&3\end{bmatrix}\times\begin{bmatrix}3&1\\3&2\end{bmatrix}$$

$$=\begin{bmatrix}19&24\\19&20\end{bmatrix}+\begin{bmatrix}15&8\\12&7\end{bmatrix}=\begin{bmatrix}34&32\\31&27\end{bmatrix}$$

$$H_{00}=H_{00}+D_{02}\times E_{20}=\begin{bmatrix}28&19\\36&31\end{bmatrix}+\begin{bmatrix}1&2\\2&5\end{bmatrix}\times\begin{bmatrix}1&3\\1&1\end{bmatrix}$$

$$=\begin{bmatrix}28&19\\36&31\end{bmatrix}+\begin{bmatrix}3&5\\7&11\end{bmatrix}=\begin{bmatrix}31&24\\43&42\end{bmatrix}$$

After the third shift

$$C_{00}=C_{00}+A_{03}\times B_{30}=\begin{bmatrix}34&32\\31&27\end{bmatrix}+\begin{bmatrix}7&2\\1&2\end{bmatrix}\times\begin{bmatrix}1&2\\2&1\end{bmatrix}$$

$$=\begin{bmatrix}34&32\\31&27\end{bmatrix}+\begin{bmatrix}11&16\\5&4\end{bmatrix}=\begin{bmatrix}45&48\\36&31\end{bmatrix}$$

$$H_{00}=H_{00}+D_{03}\times E_{30}=\begin{bmatrix}31&24\\43&42\end{bmatrix}+\begin{bmatrix}1\\4\end{bmatrix}\times\begin{bmatrix}2&3\end{bmatrix}$$

$$=\begin{bmatrix}31&24\\43&42\end{bmatrix}+\begin{bmatrix}2&3\\8&12\end{bmatrix}=\begin{bmatrix}33&27\\51&54\end{bmatrix}$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

4. BLOCK RHOTRIX OPERATIONS

Rhotrix computations involving scalar algebraic operations on the rhotrix elements can be expressed in terms of identical operations on submatrices of the original rhotrix.

In this view, an n x n rhotrix A can be regarded as a q x q array of blocks $A_{i,j}$ $(0 \leq i, j < q)$ such that each block is an (n/q) x (n/q) subrhotrix.

We perform q^3 rhotrix multiplications, each involving (n/q) x (n/q) matrices requiring $(n/q)^3$ additions and multiplications.

5. RHOTRIX-RHOTRIX MULTIPLICATION: CANNON'S ALGORITHM PATTERN

- Consider n dimensional rhotrices A_n and B_n partitioned into p blocks $A_{i,j}$ and $B_{i,j}$ $(0 \leq i, j < \sqrt{p})$ each.
- Process $P_{i,j}$ initially stores $A_{i,j}$,$B_{i,j}$ and $D_{i,j,}$ $E_{i,j}$ and computes block $C_{i,j}$ and $H_{i,j,}$of the resulting rhotrix.
- Computing subrhotrix $C_{i,j}$ requires all subrhotrices $A_{i,k,}$ $B_{k,j}$ and $D_{i,k,}$ $E_{k,j}$ for $0 \leq k < \sqrt{p}$.
- Every process in the i^{th} row requires all subrhotrices, the all-to-all broadcast can be avoided by scheduling the computations of the processes of the i^{th} row such that, at any given time, each process is using a different block $A_{i,k}$ and $D_{i,k}$ by systematically rotating these blocks among the processes after every submatrix multiplication so that every process gets a fresh $A_{i,k,}$ and $D_{i,k}$ after each rotation.
- If an identical schedule is applied to the columns of B and E, then no process holds more than one block of each matrix at any time.

6. COMMUNICATION STEPS IN CANNON'S ALGORITHM

Applying the block row-column operation using the cannon's algorithm method for rhotrices with main entries $A_{i,j}$, $B_{i,j}$ and with corresponding hearts $D_{i,j}$, $E_{i,j}$, we consider cannon's matrix multiplication algorithm from the point of view of process $P_{i,j.}$ The processes are organized into a 2-D mesh, and each process has a block of A, B, D and E needed for computing C and H respectively. Figure 9a depicts the first block multiplication step. After each block multiplication process $P_{i,j}$ sends its block of A to the process on its left and receives a new block of A from the process on its right. Similarly, it sends its block of B to the process above it and receives a new block of B from the process below it. Figure 9b, 9c and 9d shows the second, third and final block rhotrix multiplication respectively. Summing the results of all block rhotrix multiplications yields $C_{i,j}$ and $H_{i,j.}$

Blocking and Job Assignment Operation

A. First Alignment

Figure 9a: First block rhotrix multiplication step. *A, B and D, E after initial alignment*

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

B. Second Alignment

Figure 9b: Final block rhotrix multiplication step. *Sub-rhotrix location after first shift*

C. Third Alignment

Figure 9c: Final block rhotrix multiplication step.*Sub-rhotrix location after second shift*

D. Final Alignment

Figure 9d: Final block rhotrix multiplication step

First, align the blocks of A, D and B, E in such a way that each process multiplies its local submatrices:

- Shift sub matrices $A_{i,j}$ and $D_{i,j}$ to the left (with wraparound) by i steps.
- Shift sub matrices $B_{i,j}$ and $E_{i,j}$ up (with wraparound) by j steps.
- Perform local block matrix multiplication.
- Next, each block of A, D moves one step left and each block of B, E moves one step up (again with wraparound).

Perform next block multiplication; add to partial result, repeat until all the √p blocks have been multiplied.

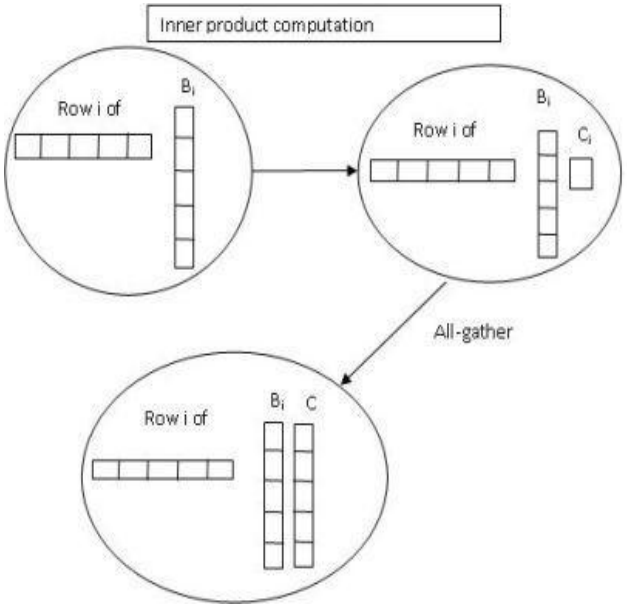


Figure 10a row-column block rhotrix multiplication requiring all-gather operation

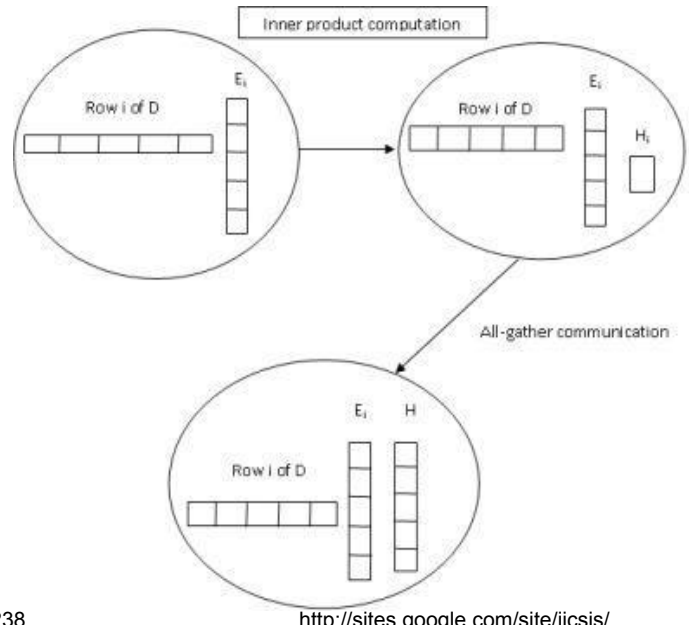


Figure 10b row-column block rhotrix multiplication

7. BLOCK RHOTRIX ALGORITHM

The block rhotrix multiplication algorithm for n x n matrices with a block size of (n/q) x (n/q)

```
procedure RHOTRIX_MULT (A, B, C, D, E, H)
begin
  for i := 0 to t - 1 do
    for j := 0 to t - 1 do
      begin
        {Initialize all elements of C_{i,j} and H_{k,j} to zero}
        C[i, j] := 0;
        H[i, j] := 0;
      for k := 0 to t - 1 do
      {Receive message from slave process (worker) i,j}
C_{i,j} := C_{i,j} + A_{i,k} x B_{k,j};
        H_{i,j} := H_{i,j} + D_{i,k} x E_{k,j};
  endfor;
 endfor;
  {Write C and H to output file}
end RHOTRIX_MULT
```

8. CONCLUSION

In this paper we discussed the basic concept of higher dimensional rhotrices multiplication and parallel algorithms for multiplying two n-dimensional rhotrices R(A) and R(B) has been proposed. The parallel rhotrix multiplication algorithms proposed is based on the conventional serial algorithm. The idea of matrix operations has been used in solving product of n x n rhotrices.

REFERENCES

[1] Ajibade, A. O.,2003, the concept of rhotrix in mathematical enrichment, International Journal of Mathematical Education in Science and Technology, 175-179.

[2] B. Sani, 2006, the row-column multiplication of high dimensional rhotrices, International Journal of Mathematical Education in Science and Technology, 1-6.

[3] B. Sani, 2006, Conversion of a rhotrix to a 'coupled matrix' , International Journal of Mathematical Education in Science and Technology, 1-7.

Abdullahi Mohd received his B.Tech. Degree from Federal University of Technology, Minna. Nigeria, in 2006. He is currently a staff in training Ahmadu Bllo University, Zaria. His area of interest is High Performance Computing.

Ezugwu E. A. received his BSc. Degree from Ahmadu Bello University, Zaria. Nigeria, in 2006. He is currently a master's degree student in same University. His area of interest is Software Engineering.

Barroon Isma'eel Ahmad received the B.Sc Degree from Usmanu Danfodiyo University, Sokoto. Nigeria, in 2006. He is currently a staff in training Ahmadu Bllo University, Zaria. His interests are Mobile Communication and Location Aware Application Development.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Clinical Data Warehouse on Insect Vector Diseases to Human of Andhra Pradesh

[1]Dr.M. Usha Rani, [2]M.Kalpana Devi, [3] Dr.D.M. Mamatha
[4] Dr.R.Seshadri , [5]Yaswanth Kumar.Avulapti

[1]Associate Professor, [2]Research Scholar, Dept. of Computer Science,
[3]Associate Professor, Dept. of Seri-Bio Sciences, SPMVV, Tirupati
[4] Director, S.V.U.Computer Center S.V.University, Tirupati
[5]Research Scholar, Dept of Computer Science ,S.V.University, Tirupati

ABSTRACT

The Widespread of Insect Vector Diseases to humans is causing substantial morbidity and economic loss to our nation. The year 2006 is likely to go down as one of the worst years in terms of public health, which has witnessed a high incidence of Insect Vector Diseases such as ***Malaria, Chikungunya, Dengue, Lymphatic Filariasis, And Japanese Encephalitis.*** *This stressed the need to track the relevant information about these diseases. The reliable and quickly retrievable clinical data on disease wise is a need of the hour with which planners can prepare their strategies to control and curb the diseases. From the aforesaid point of view this particular data warehouse (DWH) going to be handy to the planners.*

Key Words: Insect Vector diseases, Chikungunya, Malaria, Dengue, Lymphatic Filariasis & Japanese Encephalitis, Clinical data, Data warehouse

1. Introduction

The epidemic diseases are a threat to the society starting from the stone age to till date. Even though we have good past experience about epidemic diseases but the problems are not handled in a proper way. The control of these diseases involves control of three living beings and their environment viz. man-the host, mosquito-the vector and the deadly pathogen-the parasite. Since the vector and the pathogen are highly adaptable, much of the emphasis is on man i.e. bringing the awareness in public related to the insect vector diseases.

The National and International efforts over these Insect Vector Diseases control were highly successful in late 1950's and the early 60's. However, due to various reasons the control programs received setbacks all over the world and today it has come back with vengeance. Present epidemic of Chikungunya in India after a gap of 30 years, is the largest ever in the world, with over 1.3 million people affected. For other mosquito borne diseases there has been a threefold increase in Japanese Encephalitis since 2001. Malaria infects 2 million Indians annually. It is time to address the research on these lines to explore, where the system fails in combating these diseases.

2. Origin of the Research Problem

The widespread of Insect Vector diseases to human is causing substantial morbidity and economic loss to our nation. The year 2006, is likely to go down as one of the worst years in terms of public health, which has witnessed a high incidence of Insect Vector diseases such as Malaria, Chikungunya, Dengue, Japanese Encephalitis. The WHO regional office for South-East Asia has reported 1.3 million cases from 152 districts in 10 states/provinces of India, out of which 7,52,245 were from Karnataka alone. Impact on disease spread includes socio-economic aspects, clinical attendance and barriers to health care and lack of awareness to control the diseases. This stressed the need to track the relevant information, the various aspects and data about these diseases.

3. Significance of the Work

The epidemic diseases are a great threat to India and there is a need to construct the data warehouse for prevention, early detection and to take control measures. There is a need to aware the public about epidemic diseases. The information given by data warehouse is useful to the researchers, academicians, doctors, health workers and Govt. servants including common man. This data keeps us aware and forearmed to prevent such attacks in future.

4. Objectives

This work is proposed to be undertaken with the following objectives:

- Persons at the helm of affairs at central Government in general and State Government in particular are worsely in need of disease wise clinical data to equip themselves with corrective cum counter strategies. The reliable and quickly retrievable clinical data on disease wise is a need of the hour with which planners can prepare their strategies to control and curb the diseases from this point of view this particular data warehouse going to be handy to the planners.
- This data warehouse is for the future use of the researchers, academicians, Doctors, Health workers and Govt. servants including common man. This data keeps us aware and forearmed to prevent such attacks in future.
- The data warehouse and analysis reports will be made publicly available for further research.

5. Data Warehouse

Data Warehousing is a buzz-phrase that has taken the information systems' world by storm. A data warehouse (DWH) can be looked at as an "informational database" that is maintained separately from an organization's operational database. But that would fall short of the full technological implications of the DWH term. The process of transforming data into information and making it available to the user in a time bound manner to make a difference is known as data warehousing.

In order to serve the decision making process of the management the data warehouse has to supply the following primary functionality:

- The DWH is a reflection of the business rules of the enterprise – not just of a specific function or business

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

unit-as they apply to strategic decision support information.

- It is the collection point for the Integrated, Subject-Oriented strategic information that is handled by the data acquisition process.
- It is the historical store of strategic information, with the history relating to either the data or its relationships.
- It is the source of stable data regardless of how the processes may change. This requires a data model that is not influenced by the operational processes creating the data.

Additionally the data warehouse provides as functionality for the support of ad hoc queries.

5.1 The Clinical Data Warehouse

A clinical data warehouse or CDWH is a facility that houses all electronic data collected at a clinical center. For any modern clinical institute, it is necessary to separate operational data from informational data by creating a clinical data warehouse. A growing number of technologies for integrating and performing structured analyses of data from disparate sources are competing to win the day for healthcare organizations.

A CDWH is therefore a DWH tailored for the needs of users in a clinical environment, combining information from a variety of legacy health-care databases and cleansed operational data to form a centralized data repository to answer the informational needs of all clinical users.

Data warehouse in clinical context have traditionally been administrative in nature, focusing on patient billing and patient-care management, organizational aspect of hospitals that were optimized using data warehouse technology not much different than contemporary enterprises. Technology however evolved quickly and more complex areas of clinical data management could be tackled. The information technology supported collection process of clinical data has had a long history, and the promise of a new technology leveraging these collections put physicians, nurses and clinical researchers right next to the administrators on the map.

5.2 Extraction, Transformation and Loading - Three Stage Method

As data warehouse data are highly aggregated, very complex relationships are constructed from various data sources. The process that is responsible for exactly that transformations is called Extraction Transfer and Loading(ETL) process and handles getting data out of one data store[extraction], modify it [transfer], and inserting it into a different data store[loading].

Data are extracted from operational databases, legacy systems and external data sources, transformed to match the DWH schema, and loaded into the data warehouse database. Generally ETL is a complex combination of processes and technology that consumes a significant portion of the data warehouse development resources and time. Further importance is placed on the ETL process due to the fact that it is not a one-time event, but staged periodically. Typical periodicity shows in monthly, weekly, or daily updates, depending on the purpose of the data warehouse. ETL also changes as the data warehouse evolves, so ETL processes must designed for ease of modification.

Once the scope had been set, the relevant data has to be identified from the raw source data available, to formalize the approach of this task, a method which transforms data from the raw data to the source of DWH has three steps which are as follows:

- First stage data is the raw data from operational database.
- Second stage data is transformed, cleansed and normalized from stage 1 data.
- Third stage data is further transformed from stage 2, optimized for final fact data representation.

"Data stage" software is an ETL Tool selected to implement data warehouse.

5.2.1 Stage 1- Raw Data

Data collected from the **Ministry of Health and Family Welfare, Hyderabad, Andhra Pradesh**. This stage data is considered as raw source data, which are of table format. Data about all the five diseases such as Dengue Table, Malaria Table, Chikungunya Table, Japanese Encephalitis Table, and Filariasis Table are of the same format. Description of the sample table is as follows.

Dengue Table (District Name	Character
Year	Number
Total Blood Samples Collected	Number
Confirmed Cases	Number
Number of Deaths	Number)

5.2.2 Stage 2 - Refined data

Based on the tables from Stage 1 certain design decisions had to be made before any data modeling could commence. Questions like

What are the central facts?

Which are the dimensions should be focused?

Parallel to the modeling process, steps were taken to get an idea of what information could be derived from the data available. Stage one tables are used to construct text files. The description of data files are as follows:

Disease Table

No. of records: 6

We have taken five Mosquito Borne Diseases such as Malaria, Dengue, Chikungunya, Lymphatic Filariasis, and Japanese Encephalitis. Description of the table as follows.

Disease Table (S. No	Number,
Disease Id	Character,
Disease Name	Character)

(S. No=0, No disease)

District Table

No. of records : 23 (Total number of districts in Andhra Pradesh)

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Description of the table as follows.

District table (District Id Character,
District Name Character)

Case–year Table

No. of records: Number of confirmed cases of a disease in a particular year varies based on the disease and district.

We have constructed 9 text files; each contains data for a single year i.e. from 2000 to 2008.

Case Year Table (Case Id Character,
District Id Character,
Disease Id Character,
Blood Samples
Collected or Not Logical,
Year Number,
Disease Status Logical)

Later we combine all the 9 text files into a single text file called CASE_ENTIRE_YEAR text file.

Case–District-2009 Table

Total **No. of Records: 33601**

We have constructed 23 text files for 23 districts each contain data for the year 2009. (These tables are for the current year updation.)

Case District 2009(Case Id Character,
Disease Id Character,
Blood Samples Logical,
Year Number,
Disease Status Logical)

Later we combine all the 23 districts text files into a single text file called CASE_HISTORY text file.

Death Table

No. of records: 1200

Death table (Case Id Character,
District Id Character,
Disease Id Character,
Blood Samples
Collected or not Logical,
Year Number,
Disease Status Logical,
Death Id Character)

5.2.3 Stage3-LoadingStage(Clinical Warehouse creation)

Data from all the text files was extracted and stored in an Oracle file, while transforming primary keys are to be specified in the oracle table.

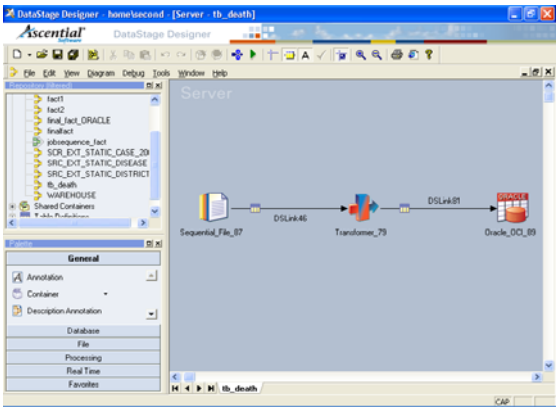
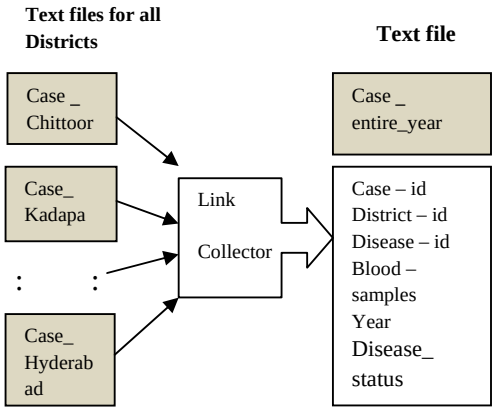


Fig: 1

Data from all the Case tables from 23 districts for a single year (for example 2009 data) are combined using link collector into another sequential file by using Round Robin Algorithm.



Case_2009 Job

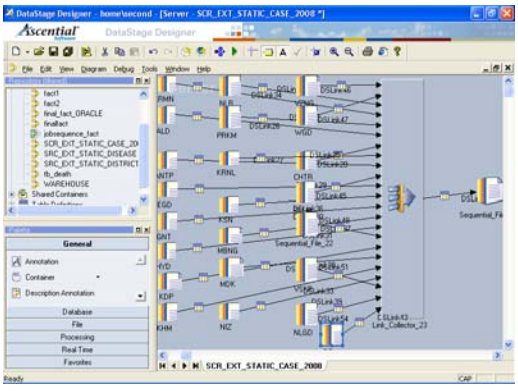
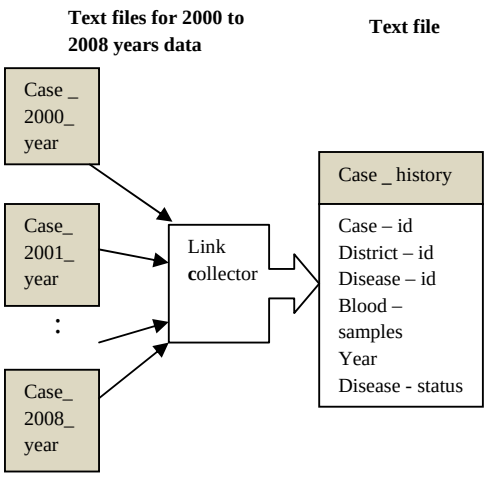


Fig: 2

The resulted sequential table is then transformed to Oracle table with the same attributes.

Data from all the Case-year tables for 9 years (2000 to 2008) are combined using link collector into another sequential file by using Round Robin Algorithm.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500



Case_History Job

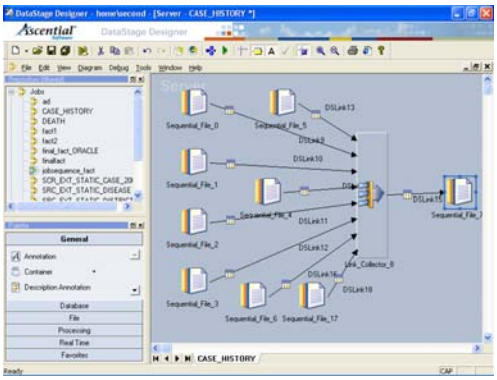


Fig: 3

The resulted sequential table is then transformed to Oracle table with the same attributes.

Fact 1 Table

Hash file 1: To find the death count we apply query on the Death table

Query applied:

Select Count (*), District_id, Disease_id, Year
From Death_table
Group by District_id, Disease_id, Year;

Hash file 2: To find case count we apply query on the Case table

Query applied :

Select Count (*), District_id, Disease_id, Year
From Case_table
Group by District_id, Disease_id, Year;

Hash file 3: To find total blood samples collected we apply query on the Case table

Query applied:

Select Count (*), District_id, Year
From Case_table
Where blood_samples = 'y'
Group by District_id, Year;

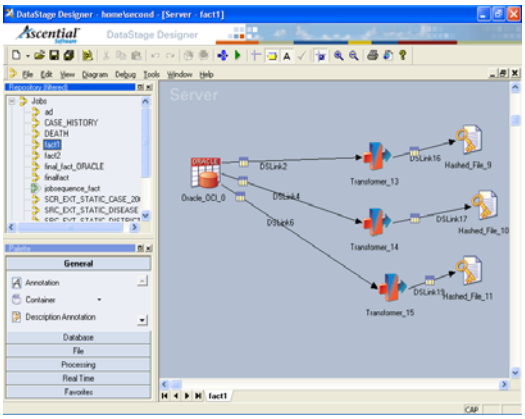


Fig: 4

Fact 2 Table

Hash file 4: To find Distinct District_id, Disease_id from Case table

Query applied:

Select distinct District_id, Disease_id, Year

From Case_table;

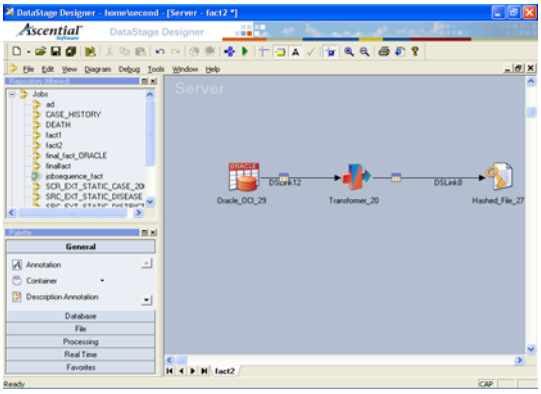


Fig: 5

All the 4 Hash tables, Oracle District table and Disease table are combined and transformed to create final fact table. In the transformation process we apply the following queries on the Oracle Disease table and District table to include the names of the Disease and Districts in the final fact table.

Query applied on Disease table:

Select Disease_tb. Disease_id, Disease_tb.Disease _name
From pro.Disease_tb Disease_tb
Where Disease_tb. Disease_id=:1;

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Query applied on District table:

Select District_tb.District_id, District_tb. District _name
From pro.District_tb, District_tb
Where District_tb.District_id =:1;

Final Fact Job:

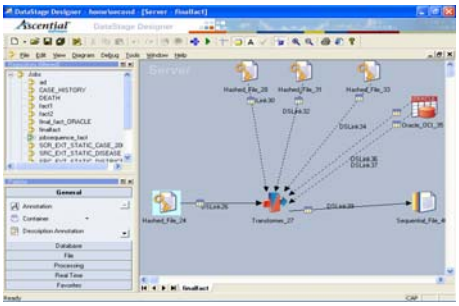
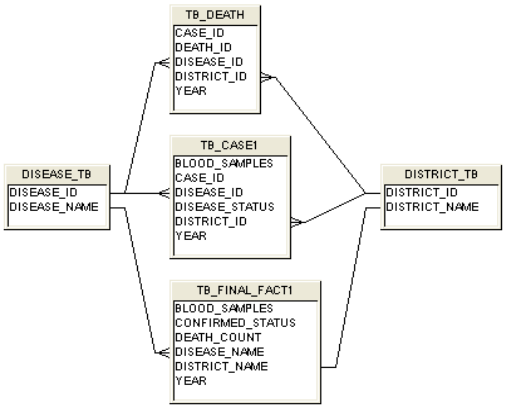


Fig: 6

5.2.4 Data Model:

Data modeling technique we used is Star schema. The advantage of Star schema is that it is easy to understand, easy to define hierarchies, reduces the number of physical joins, and requires low maintenance and simple Meta data. The actual data model for this data warehouse is as follows.



5.3 Reporting Tool:

Data Stage ETL Tool is used to create data warehouse .The final fact table produced from ETL Tool will be given to the Reporting Tool, which will produce Reports . Reporting Tool we used is **Business Objects**.

Requested sample query:

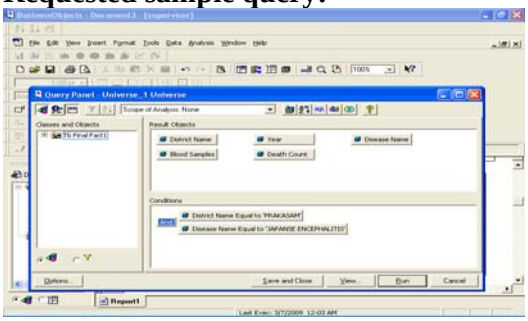


Fig: 7

Sample Report produced :

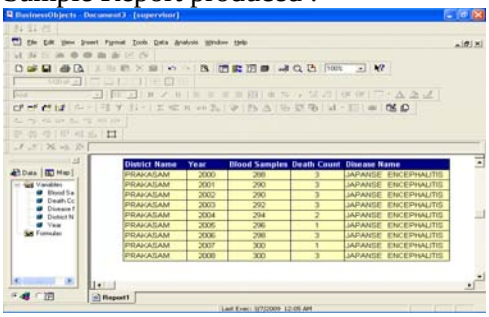


Fig: 8

Conclusions

Modern tools now come in handy to address the issues on the disease surveillance, control, monitoring and evaluation, where should health care centers to be situated and what services should they offer. Monitoring and evaluation are an essential part of the health programme as well as other programmes related to development. Hence, there is a need to sensitize the public about epidemic diseases. This stressed the need to construct the data warehouse for prevention, early detection and to take control measures. This data keeps us aware and forearmed to prevent such attacks in future.

The work is concentrated towards to build the data warehouse. Due to the time limitation, the current history file is constructed only based on the data from 2000 to 2009. This data to be extend regularly with the availability of the next year data. This data warehouse is for the future use of the researchers, academicians, Doctors, Health workers and Govt. servants.

References

[1] Sid Adelman, Larissa Moss, “Data Warehouse Project Management”, 2000.
[2] Codd E.F., “The Clinical Data Warehousing”, 2001.
[3] www.etl-tools.info/en/bi/etl-process.htm
[4] Jonathan G. Geiger, “The Data Warehouse Model”, 2000.
[5] Michael Haisten, “The real-time data warehouse: the next stage in Data warehouse evolution”, 1999.
[6] Harry Singh, “Interactive Data Warehousing”, PHI, 1994.
[7] William Inmon, “Building the Data Warehouse”, 2nd ed, John Wiley, New York, 1996.
[8] “Clinical Data Warehouse Functionality” Peter Villiers, SAS Institute Inc., Cary, NC
[9] Laura Hadley, “Developing Data warehousing Architecture”, 2001.
[10] Steven R. Meyer, “Which ETL Tool is Right for You?”, DM Review, Mar 2001.
[11] Nigel Pendse, “What is OLAP? The Codd Rules and Features”, 2008.
[12] National Center for Infectious Diseases, Division of Parasitic Disease.
[13] DJ. Power, “A Brief History of Decision Support Systems”, 2002.
[14] Ralph Kimball, “The Data Warehouse Toolkit”, 1994.
[15] Warren Thornthwaite, “Understanding Data Warehouse Architecture requirements”, 2000.
[16] Usha Rani M., Jyothi S., Rama Sree R. J., “Data Warehousing and Data Mining”, Ikon Books, 2009.

.Development of an Expert System for Diagnosis and appropriate Medical Prescription of Heart Disease Using SVM and RBF

Shaikh Abdul Hannan[1],
[1]Lecturer, Department of Computer Science, Vivekanand College, Aurangabad, Maharashtra, India,
hannan_7us@yahoo.com

V. D. Bhagile[2],
[2]Lecturer, Department of Computer Science, Sayali College, Aurangabad, Maharashtra, India.
v_shinde68@yahoo.co.in

R. R. Manza[3],
[3]Lecturer, Dept. of Comp. Science and I.T. Dr. B.A.M.University, Aurangabad, Maharashtra
manzaramesh@gmail.com

R. J. Ramteke[4]
[4]Reader, Dept. of Computer Science, North Maharashtra University, Jalgaon, Maharashtra,
rakeshramteke@yahoo.co.in

***Abstract*---**

Human disease diagnosis is a complicated process and requires high level of expertise. Any attempt of developing a expert system dealing with human disease diagnosis has to overcome various difficulties. This paper describes a project work aiming to develop a expert system for diagnosing of heart disease using neural network technique. Now a days neural network are being used successfully in an increasing number of application areas. This project includes the detailed information about patient and preprocessing was done. The Support Vector Machine (SVM) and Radial Basis Function (RBF) have been applied over the data for the experiment. This research project focuses on the research and development of a web-based clinical tool designed to improve the quality of the exchange of health information between health care professionals and patients. Practitioners can also use this web-based tool to corroborate diagnosis. The proposed system is experimented on various scenarios in order to evaluate it's performance. In all the cases, proposed system exhibits satisfactory results.

Keywords – Neural Network, Support Vector Machine (SVM), RBF (Radial Basis Function), Symptoms, Medicine.

I. INTRODUCTION

Expert systems(ES) are artificial intelligence-base computer programs that have received a great deal of attention during years. These programs have been used to solve an impressive array of problems in a variety of fields. The part of the expert system that stores the knowledge is called the knowledge base. The part that holds the specifics of the to-be-solved problem is call the global database. The part that applies the knowledge to the problem is called the inference engine. Expert systems typically have friendly user interface to enable inexperienced users to specify problems for the system to solve and to understand the system's conclusion.

Although, there are many computer based diagnosis systems are developed for medicine. [1-3], however the number of expert system for human heart diagnosis domains are still very few [4].

The Need for Expert Systems

Expert systems are necessitated by the limitations associated with conventional human decision-making processes, including:

1. Human expertise is very scarce.
2. Humans get tired from physical or mental workload.
3. Humans forget crucial details of a problem.
4. Humans are inconsistent in their day-to-day decisions.
5. Humans have limited working memory.
6. Humans are unable to comprehend large amounts of data quickly.
7. Humans are unable to retain large amounts of data in memory.
8. Humans are slow in recalling information stored in memory.
9. Humans are subject to deliberate or inadvertent bias in their actions.
10. Humans can deliberately avoid decision responsibilities.
11. Humans lie, hide, and die.

The proposed methodology uses neural network for classifier. The performance of proposed methodology was evaluated with two different neural network techniques. Moreover, we compared our result with Support Vector Machine and Radial Basis Function with original medicines provided by the doctor. We obtain 97% accuracy from the experiments made on the data set containing 300 samples. This classification is the highest so for with our data. The paper is organized as following, in Section 2, a brief overview on previous related works and in section 3, introduction of Support Vector Machine and Radial Basis Function is described. Section 4, the proposed methodology and preparing Data for underlying neural network. Section 5, Experimental

1. **Shaikh Abdul Hannan**-Working as lecturer in Vivekanand College, Aurangabad and persuing Ph.D. from North Maharashtra University, Jalgaon (hannan_7us@yahoo.com)
2. **V. D. Bhagile**-Working as Principal in Sayali College, Aurangabad. (v_shinde68@yahoo.com)
3. **R. R. Manza** - Working as Lecturer in Department of Computer Science and Information Technology, Dr. B. A. M. University, Aurangabad. (manzaramesh@gmail.com)
4. **R. J. Ramteke**-Working as a reader in Department of Computer Science, North Maharashtra University,Jalgaon. (rakeshramteke@yahoo.co.in)

.analysis and how the coding is done with patients as well as medicine data is described. Section 6, Discussion and result of first five patients medicine given by the expert system and is compared with the original medicine. Finally, we concluded this paper in Section 7.

II. BACKGROUND

A. Related Works

Up to now, various classification algorithms have been employed on Turkoglu's valvular heart disease data set and high classification accuracies have been reported in the last decade [5-10]. Turkoglu's valvular heart disease data set was obtained from Firat Medical Center. A detailed description for the data set will be given in the next section.

The valvular heart disease data set was firstly utilized in [5] where Turkoglu et al. fulfilled an expert diagnosis system which uses backpropagation artificial neural networks (BP-ANN) classifier. The performance evaluation of the realized system was evaluated by classification accuracy and the correct classification rate was about 94% for normal subjects and 95.9% for abnormal subjects. Later, Turkoglu et al. suggested an intelligent system for detection of heart valve disease based on wavelet packet neural networks (WPNN) [6]. The reported correct classification rate was about 94% for abnormal and normal subjects. Recently, Comak et al. investigated the use of least-square support vector machines (LS-SVM) classifier for improving the performance of the Turkoglu's proposal [7]. They intended to realize a comparative study. Classification rates of the examined classifiers were evaluated by ROC curves based on the terms of sensitivity and specificity. The application results showed that according to the ROC curves, the LS-SVM classifier performance was comparable with ANN, but the training time of LS-SVM is shorter than that of the ANN and it can always converge the same solution while ANN cannot. According to these results, LS-SVM's training time is about 13 times shorter than ANN's training time. This is an important difference. Because, LS-SVMs are trained only depending on support vectors, not by whole training data set. In addition, LS-SVM can overcome the overfitting much successfully than ANN.

More recently, Uguz et al. performed a biomedical system based on Hidden Markov Model for clinical diagnosis and recognition of heart valve disorders [8]. The fulfilled methodology was also used the database of Turkoglu et al. In the presented study, continuous HMM (CHMM) classifier system was used. Single Gaussian model was preferred to determine emission probability. The methodology was composed of two stages. At the first stage, the initial values of average and standard deviation were calculated by separating observation symbols into equal segments according to the state number and using observation symbols appropriate to each segment. At the second stage, the initial values of average and standard deviation were calculated by separating observation symbols into the clusters (FCM or K-means algorithms) that have equal number of states and using observation symbols appropriate to the separated clusters. The implementations of the experimental studies were carried out on three different classification systems such as CHMM, FCM–K-means/CHMM and ANN. These experimental results were obtained for specificity and sensitivity rates 92% and 94% for CHMM, 92% and 97.26% for FCM–K-means/CHMM), respectively. Finally, Sengur et al. investigated the use of principal component analysis (PCA), artificial immune system (AIS) and fuzzy *k*-NN to determine the normal and abnormal heart valves from the Doppler heart sounds [9]. For reducing the complexity, PCA was used. In the classification stage, AIS and fuzzy *k*-NN were used. To evaluate the performance of the methodology, a comparative study was realized by using a data set containing 215 samples. The validation of the method was measured by using the sensitivity and specificity parameters; 95.9% sensitivity and 96% specificity rate was obtained. Sengur et al. also investigated the use of Linear Discriminant Analysis (LDA) and Adaptive neuro-fuzzy inference system (ANFIS) for clinical diagnosis and recognition of heart valve disorders [10]. The validation of the method is measured by using the sensitivity and specificity parameters. 95.9% sensitivity and 94% specificity rate was obtained.

III. ORGANIZATION TO EXPERT SYSTEM

To diagnose the heart failure cause different popular methods used are MRI, Doppler and Expert System. MRI can provide clear three dimensional images of the heart. Doppler technique has gained much more interest since Satomura first demonstrated the application of the Doppler effect to the measurement of blood velocity in 1959[11]. However the factor such as calcified disease or obesity often result in a diagnostically unsatisfactory. Doppler techniques assessment and therefore, it is sometimes necessary to assess the spectrogram of the Doppler Shift signals to elucidate m the degree of the disease [12]. Many studies have been implemented the classify Doppler signals in the pattern recognition field [13-14].

Expert system is a intelligent program which olds the accumulated knowledge of one or more domain experts. There are many types of expert system currently exist. MYCIN expert system is used in medical field for diagnosis of blood disorders. DESIGN ADVISOR is another expert system used in processor chip design to give advice to designer about component placement, minimizing chip size etc. PUFF expert system is also used in medical system for diagnosis of respiratory condition of patient. PROSPECTOR expert system is used by geologists to identify sites for drilling or mining. DENDRAL expert system is used to identify the structure of chemical compounds. LITHIAN expert system gives advice to archaeologists to examiner stone tools. Expert system having three main components knowledge base, inference engine and user interact. Knowledge base is the collection of facts and rules which describe all the knowledge about problem domain. The inference engine is used to choose the appropriate facts and rules to apply during user query. Where as user interface takes the user query in a readable form and passes it to the inference engine. It then displays the result to the user. Based on much more useful but it has some limitation like limited domain, no current updation, no system self learning, no common sense, expert needed to setup and maintain. But even though it is used

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

.world wide because they are not always available, can be used anytime anywhere, human experts are not 100% reliable or consistent. Expert may not good for explanation of decisions and cost effective. While using the expert system some legal and ethical issues we need to follow to set the responsibility.[15]

A. *Support Vector Machine :*

Support Vector Machines(SVMs) are a state of the art pattern recognition techniques whose foundation stem from statistical learning theory. However, the scope of SVMs goes beyond pattern recognition because they can also handle two more learning problems i.e. regression estimation and density estimation. An SVM is a general algorithm based on guaranteed risk bounds of statistical learning theory i.e. the so called structural risk minimization principle. It is a learning machine capable of implementing s set of functions that approximate best the supervisor's response with an expected risk bounded by the sum of the empirical risk and Vapnik – Chevonenkis (VC) confidence. Recent advances in statistics, generalization theory, computational learning theory, machine learning and complexity have provided new guidelines and deep insights into the general characteristics and nature of the model building/learning/fitting process [16]. Some researchers have pointed out that statistical and machine learning models are not all that different conceptually [17,18]. Many of the new computational and machine learning methods generalize the idea of parameter estimation in statistics. Among these new methods, Support Vector Machines have attracted most interest in the last few years.

Support vector machine (SVM) is a novel learning machine introduced first by Vapnik [19]. It is based on the Structural Risk Minimization principle from computational learning theory. Hearst et al. [20] positioned the SVM algorithm at the intersection of learning theory and practice: ''it contains a large class of neural nets, radial basis function (RBF) nets, and polynomial classifiers as special cases. Yet it is simple enough to be analyzed mathematically, because it can be shown to correspond to a linear method in a high dimensional feature space nonlinearly related to input space.'' In this sense, support vector machines can be a good candidate for combining the strengths of more theory-driven and easy to be analyzed conventional statistical methods and more data-driven, distribution free and robust machine learning methods.

In the last few years, there have been substantial developments in different aspects of support vector machine. These aspects include theoretical understanding, algorithmic strategies for implementation and reallife applications. S VM has yielded excellent generalization performance on a wide range of problems including bioinformatics [21,22,23], text categorization [24], image detection [25], etc. These application domains typically have involved high-dimensional input space, and the good performance is also related to the fact that SVM's learning ability can be independent of the dimensionality of the feature space.

The SVM approach has been applied in several financial applications recently, mainly in the area of time series prediction and classification [26,27]. A recent study closely related to our work investigated the use of the SVM approach to select bankruptcy predictors. They reported that SVM was competitive and outperformed other classifiers (including neural networks and linear discriminant classifier) in terms of generalization performance [28]. In this study, we are interested in evaluating the performance of the SVM approach in the domain of heart disease in comparison with that of Radial Basis Function in neural networks.

Let us define labeled training examples [xi, yi], an input vector $x_i \in R^n$ a class value $y_i \in \{-1,1\}, i = 1,.....,l$ For the linearly separable case, the decision rules defined by an optimal hyperplane separating the binary decision classes is given as the following equation in terms of the support vectors

$$Y = \text{sign}\left(\sum_{i=1}^{N} y_i \alpha_i (\mathbf{x} \cdot \mathbf{x}_i) + b\right) \quad (1)$$

where Y is the outcome, yi is the class value of the training example xi, and . represents the inner product. The vector x =(x1,x2,.,xn) corresponds to an input and the vectors xi, i=1,.,N, are the support vectors. In Eq. (1), b and α_i are parameters that determine the hyperplane.

For the non-linearly separable case, a high-dimensional version of Eq. (1) is given as follows:

$$Y = \text{sign}\left(\sum_{i=1}^{N} y_i \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b\right) \quad (2)$$

The function K(x,xi) is defined as the kernel function for generating the inner products to construct machines with different types of non-linear decision surfaces in the input space.

B. *Radial Basis Function (RBF)*

Radial basis function (RBF) networks were introduced into the neural network literature by Broomhead and Lowe [29]. The RBF network is similar to a general feed-forward neural network trained using the back-propagation scheme. It has three layers of neurons, namely input, hidden and output. However it uses only one hidden layer, each neuron in which operates as the Gaussian transfer function, as against the sigmoid function [30]

Mathematically, the output y of an RBF network corresponding to input x is computed by the equation;

$$y = f(x) = \sum_{i=1}^{n} w_i \varphi \| x - c_i \| + w_0 \quad \text{(Eq.A.1)}$$

Where w_i is the connection weight between the ith hidden neuron and output neuron; w_0 the bias. $\varphi||x\text{-}c_i||$ indicates a RBF which is normally Gaussian having following expression;

$$\varphi \| x - c_i \| = -\exp\left(-\sum_{i=1}^{n} \frac{\| x_i - c_i \|^2}{2\sigma_i^2}\right) \quad \text{(Eq. A.2)}$$

Where c_i are centers of the receptive field; and σ_i the widths of the Gaussian function which indicates the selectivity of a neuron.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

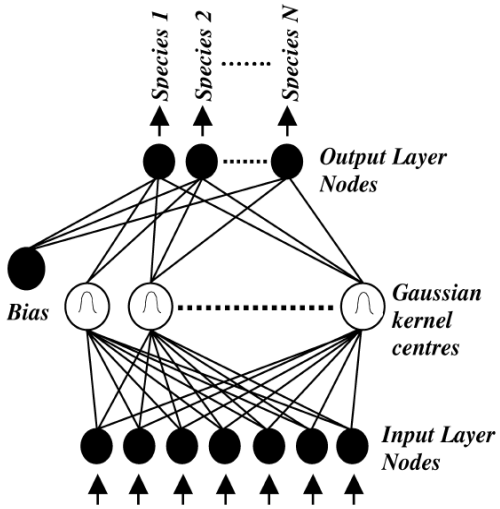
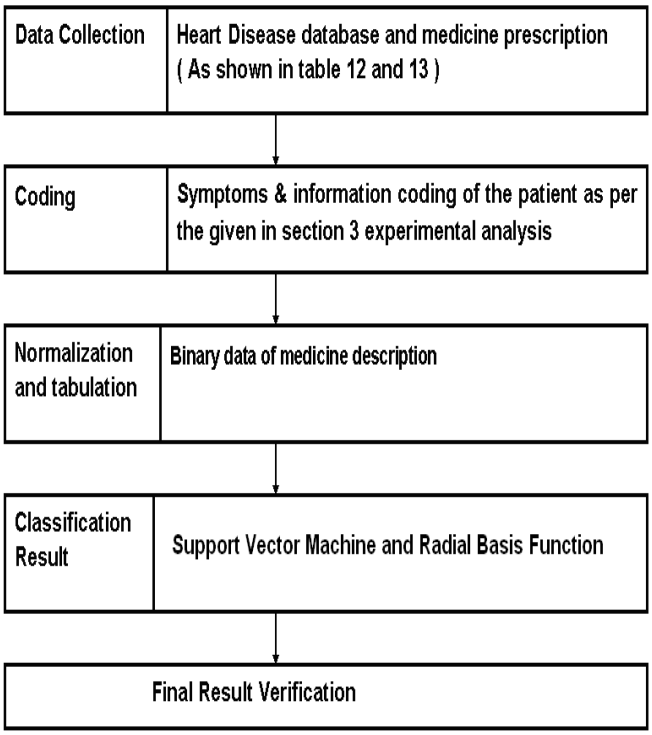


Figure 2. Schematic diagram of RBF

The major task of RBF network design is to determine centers. The easiest way to do so is to choose the centers randomly from the training set. Another approach is to use the k-means technique consisting of clustering the input training set into groups and choose the mean of each group as the center. Also, the centers can be treated as a network parameter along with w_i and adjusted through error-correction training. After the center is determined, the connection weights w_i between the hidden layer and output layer can be obtained through ordinary back-propagation-based training [30].

IV. METHODOLOGY

A. Proposed methodology and implementation of with FFBP and RBF



B. *Preparing Data for underlying neural network*:

The data is collected from daily OPD session while doctor examining the patients. The symptoms and information about patients details like Previous History(p1), Present History(p2), Personnel History(p3), Physical Examination(p4), Cardio Vascular System(CVS), Respiratory Rate(RS), Per Abdomen(PA), Central Nervous system(CVS), ECG and Blood Investigation(BI). The main point is ECG from which the patient can easily diagnose whether the patient is having heart problem or not.

All 300 patients data collected regarding heart disease and the data are prepared in different Excel Sheets which contains codes of each individual disease, history and symptoms. In one excel file 13 sub-sheets are taken for each field of information such as for Previous History (p1), for Present History the second sub-sheet and the name is given (P2), for Personnel History (P3) the third sub-sheet is taken, like this the data collection has 13 different sub-sheets for different fields. All the fields are taken under the supervision of the Cardiologist

The code is given to each symptoms, physical examination parameter or diseases in each sub-sheet for experimental work. On this data some pre-processing i.e. normalization, coding and decoding methods are applied for the expected output.

In table 1, the Previous History (P1) has represented with 1 to 18 different diseases of total 300 heart patients and represents the codes respectively from 1 to 18. The code 1 which represents Hypertension, Code 2 represents Diabetes Mallitus like this it contains 18 different diseases. Some of them are as shown in Table 1.

Code	Name of Disease
1	Hypertension
2	Diabetes Mallitus
3	TB
4	Bronchial Asthama
5	Hyperthyroidism

Table 1 : Previous History of Patients

In table 2, Present History (P2) and the symptoms present in P2 are represented by Codes. The Code 1 which represents Chest Pain/Discomfort, Code 2 represents Retrosternal Pain like this it contains 29 different symptoms. Some of the symptoms are shown in table 2.

Code	Symptoms
1	Chest Pain/Discomfort
2	Retrosternal Pain
3	Palpitation
4	Breathlessness
5	sweating

Table 2 : Present History of patients

In table 3, Personnel History (P3) and the information present in P3 are represented by codes for different bad habits. The Code 1 which represents Smoking, Code 2 represents Tobacco like this 4 different bad habits are taken and specified by 1 to 4 codes. Some of the personnel history parameters are given below.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Code	Personnel History
1	Smoking
2	Tobacco
3	Alcohol
4	Nil

Table 3 : Personnel History

In table 4, Physical Examination (P4) and the information present in P4 are represented by codes for different physical parameters. The Code 1 which represents Consciousness, Code 2 represents Orientation like these 25 different physical parameters and specified by 1 to 25 codes for each parameter. Some are as shown below in table 4.

Code	Physical Examination
1	Altered Consciousness
2	Orientation
3	Dyspnoea
4	Fever
5	Low Pulse Rate

Table 4 : Physical Examination

In table 5, Cardio Vascular System (CVS) and the information present in CVS are represented by codes for different symptoms. The Code 1 which represents Heart Sound, Code 2 represents Normal Heart Rate like this 8 different symptoms and specified by 1 to 8 codes for each symptom. Some are as shown below in table 5.

Code	Symptoms
1	Heart Sounds
2	Normal Heart Rate
3	Tachycardia
4	Bradycardia
5	Regular Heart Rhythm

Table 5 : Cardio Vascular System

In table 6, Respiratory System (RS) and the information present in RS are represented by codes for different symptoms. The Code 1 which represents Breath Sound preserved, Code 2 represents Breath Sound Reduced like this 5 different symptoms are found and specified as shown in table 6.

Code	Symptoms
1	Breath Sounds Preserved
2	Breath Sound Reduced
3	Basal Crepts
4	No Abnormality Detected (NAD)
5	Ranchi

Table 6 : Respiratory System

In table 7, Per-Abdomen (PA) and the information present in PA are represented by codes for different symptoms. The Code 1 which represents Liver (Hepatomegaly), Code 2 represents Spleen (Splenomegaly) like these 6 different symptoms have found and specified by 1 to 6 codes for each symptom. Some are as shown below in table 7.

Code	Symptoms
1	Liver(Hepatomegaly)
2	Spleen (Splenomegaly)
3	Free Fluid Present
4	Abdominal Distension
5	No Abnormality Detected (NAD)

Table 7 : Per Abdomen

In table 7, Central Nervous System (CNS) and the information present in CNS are represented by codes for different symptoms. The Code 1 which represents Consciousness, Code 2 represents Orientation like this 5 different symptoms are found and specified by 1 to 5 codes for each symptom. Some are as shown below in table 8.

Code	Symptoms
1	Consciousness
2	Orientation
3	Focal Deficit
4	No Abnormality Detected (NAD)
5	Restlessness

Table 8 : Central Nervous System

In table 8, Electro Cardio Gram (ECG) and the information present in ECG are represented through codes for different finding which points to different problems of heart. The Code 1 which represents ST Elevation, Code 2 represents Anterior Wall like this 21 different heart findings are found and specified by 1 to 21 codes for each finding. Some are as shown below in table 9.

Code	ECG Point
1	ST Elevation
2	Anterior Wall
3	Antero Septal
4	Inferior
5	Infero Posterior

Table 9 : Electro Cardio Gram (ECG)

In table 10, Blood Investigation (BI) and the information present in BI are represented through codes for blood investigation. The Code 1 which represents Cardiac Enzymes (High), Code 2 represents Blood Sugar Test like this 24 different investigations has found and specified by 1 to 24 codes for each investigation in all patient. Some are as shown below in table 10.

Code	Symptoms
6	Lipid Profile normal
7	Lipid Profile Abnormal
8	Complete Blood Count Normal
9	Leucocytosis
10	Anaemia

Table 10 : Blood Investigation

In table 11, all the medicines names along with their codes i.e. MID which are prescribed by the doctor to the patients. The medicine sheet contains 52 different medicines which are prescribed by the doctor in different 300 stages. Some are as shown below in table 11.

Code	Medicine Name
1	Alprazolam
2	Amlodepine
3	Aspirin
4	Atenolol
5	Atorvastatin

Table 11 : Medicine Names

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

In table 12, all Patients information such as previous history(P1), P2(Present History), P3(personnel History), P4(Physical Examination), CVS(Cardio Vascular System), RS(Respiratory System), PA(Per Abdomen), CNS(Central Nervous System), ECG(Electrocardiography) and BI(Blood Investigation) which contains all the represented codes that are present in the individual patients.

Sr. No.	Patient Name	Symptoms and Findings										
		Age	P1	P2	P3	P4	CVS	RS	PA	CNS	ECG	BT
1	A	55M	2	1,2,5,13	4	7,10	8	4	5	4	1,3	14
2	B	58 M	2	1,2,8	2	7,8,13,14	8	4	5	4	2	7
3	C	60M	8	5,7,13	4	1,6,12	8	4	5	4	9	14
4	D	60M	1,2	4,5	4	1,2,7,13,14	3,5	3	5	4	12	4
5	E	56F	1	15,16	4	6,9,12	8	4	5	4	10	2

Table 12 : collection of different details of the individual Heart Patients

In table 13, different 52 medicines were used by the doctor on total 300 patients. All the medicines are prescribed by the doctor. In this table the medicines codes are used as the description given in the table 11.

Sr. No.	Patient Name	MID 1	MID 2	MID 3	MID 4	MID 5	MID 6	MID 7	MID 8	MID 9	MID 10	MID 11	MID 12	MID 13
1	A	2	3	5	6	14	17	19	21	23	25	26	27,29	36
2	B	2	3	5	6	14	16	17	21	23	25	26	27	28
3	C	1	5	6	14	25								
4	D	3	5	7	10	11	13	14	17	19	30			
5	E	5	14	15	19									

Table 13 : All the Medicine codes provided by the doctor to the individual patients.

V. EXPERIMENTAL ANALYSIS

For further training of neural network process the proposed information is coded in binary form (0 or 1). If the symptom is present in the patients at particular position at that point it is defined by one (1) and if the symptoms or disease is not present at that position it is placed by Zero (0). Suppose for example in the field P2 (present history) there are total 29 symptoms present and the patient no 1 is having the symptom 1,2,5 and 13 so these locations are defined by 1 (one) and all other symptoms are 0 (zero). In such a way all the fields are defined. All the parameter that we consider in medical prescription like Sr. No., ,age , P1, P2,P3,P4,CVS, RS, PA, CNS, ECG and BT are converted in binary number where this is used in neural network for train the neurons for achieving better result.

The individual data of the patient no 1 is defined in binary form as :

Sr No **Age** **P1** **P2** **P3**

00000001 0110111 01000000000000000 11001000000010000000000000000 0001

P4 **CVS** **RS** **PA** **CNS** **ECG**

00000010010000000000000000 00000001 00010 000010 00010 10100000000000000000

BT

00000000000000010000000000;

Symptoms and Information Coding of the patient 1.

Using this sequence of binary format we were not getting appropriate result. Therefore we have change the order of fields as per suggestion of the doctor because the doctors are prescribing the medicines on the basis of the ECG and blood investigation. So the order of ECG is changed from field no. 9 to field no. 1 and after ECG we have taken Blood Investigation and rest of the fields are same and at last age is placed. Due of reshuffling of the fields we got satisfactory result upto 97%.

For this expert system total 52 different medicine are prescribed by the Doctor and if the medicine is present at that position it is defined by one (1) and if it is absent at that position it is defined by Zero(0). Similarly for patient one the prescribed medicine are defined as :

0 1 1 0 1 1 0 0 0 0 0 0 0 1 0 0 1 0 1 0 1 0 1 0 1 1 1 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0;

Medicine Coding of the patient 1.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

MID	Patient from A to I								
	A	B	C	D	E	F	G	H	I
1	1	1	1	1	1	1	1	1	0
2	0	0	0	0	1	0	1	1	1
3	1	1	1	1	1	0	1	1	1
4	0	0	0	0	0	0	1	0	0
5	1	1	1	1	1	1	1	1	1
6	1	1	1	1	0	1	1	0	1
7	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0
11	0	0	1	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0
13	0	0	0	1	0	0	0	0	0
14	0	0	1	1	1	0	1	1	0
15	0	0	0	0	0	0	0	0	0
16	1	1	0	0	0	1	0	0	1
17	1	1	0	1	0	1	1	0	1
18	1	1	0	0	0	0	0	0	0
19	1	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0
21	1	1	1	1	0	1	0	0	0
22	0	0	1	1	0	0	0	0	0
23	1	1	1	1	0	1	0	0	0
24	0	0	1	0	0	0	0	0	0
25	1	1	1	1	0	1	0	0	1
26	1	1	1	1	0	0	0	0	0
27	1	1	1	1	0	0	0	0	0
28	1	1	0	1	0	0	0	0	0
29	1	1	0	0	0	1	0	0	1
30	0	0	0	0	0	0	0	0	0
31	0	0	0	0	0	1	0	0	0
32	0	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0
36	0	0	1	0	0	1	0	0	0
37	0	0	0	0	0	0	0	0	0
38	0	0	0	0	0	0	0	0	0
39	0	0	0	0	0	0	0	0	0
40	0	0	0	0	0	0	0	0	0
41	0	0	0	0	0	0	0	0	0
42	0	0	0	0	0	0	0	0	0
43	0	0	0	0	0	0	0	0	0
44	0	0	0	0	0	0	0	0	0
45	0	0	0	0	0	0	0	0	0
46	0	0	0	0	0	0	0	0	0
47	0	0	0	0	0	0	0	0	0
48	0	0	0	0	0	0	0	0	0
49	0	0	0	0	0	0	0	0	0
50	0	0	0	0	0	0	0	0	0
51	0	0	0	0	0	0	0	0	0
52	0	0	0	0	0	0	0	0	0

Table 14 shows the Result of the first nine patient after training using SVM

MID	Patient from A to I								
	A	B	C	D	E	F	G	H	I
1	0	1	1	1	0	1	1	1	1
2	0	0	0	0	0	0	0	0	0
3	1	1	1	1	1	0	1	1	1
4	0	0	0	0	0	0	0	0	0
5	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1
7	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0
11	0	1	0	1	0	0	0	0	1
12	0	0	0	0	0	0	0	0	0
13	0	0	0	1	0	0	0	0	0
14	1	1	1	1	1	0	1	1	1
15	0	0	0	0	1	0	0	0	0
16	1	1	0	0	0	1	0	0	0
17	1	1	0	1	0	1	1	1	1
18	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0
21	1	1	1	1	0	1	1	1	1
22	0	1	0	0	0	0	0	0	0
23	1	1	0	0	0	1	0	0	0
24	0	1	0	0	0	0	0	0	0
25	1	1	1	1	0	1	1	1	1
26	1	1	0	0	0	0	0	0	0
27	1	1	0	0	0	0	0	0	0
28	1	0	0	0	0	0	0	0	0
29	1	0	0	0	0	1	0	0	0
30	0	0	0	1	0	0	0	0	0
31	0	0	0	0	0	1	0	0	0
32	0	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0
36	0	1	0	0	0	1	0	0	0
37	0	0	0	0	0	0	0	0	0
38	0	0	0	0	0	0	0	0	0
39	0	0	0	0	0	0	0	0	0
40	0	0	0	0	0	0	0	0	0
41	0	0	0	0	0	0	0	0	0
42	0	0	0	0	0	0	0	0	0
43	0	0	0	0	0	0	0	0	0
44	0	0	0	0	0	0	0	0	0
45	0	0	0	0	0	0	0	0	0
46	0	0	0	0	0	0	0	0	0
47	0	0	0	0	0	0	0	0	0
48	0	0	0	0	0	0	0	0	0
49	0	0	0	0	0	0	0	0	0
50	0	0	0	0	0	0	0	0	0
51	0	0	0	0	0	0	0	0	0
52	0	0	0	0	0	0	0	0	0

Table 15 shows Result of the first nine patient after training using RBF

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

MID	Patient from A to I								
	A	B	C	D	E	F	G	H	I
1	1	0	1	0	0	1	1	1	0
2	0	1	0	0	0	0	0	0	0
3	1	1	0	1	1	1	1	1	0
4	0	0	0	0	0	0	0	0	0
5	1	1	1	1	0	1	1	1	0
6	1	1	1	0	0	1	1	0	0
7	0	0	0	1	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0
10	0	0	0	1	0	0	0	0	0
11	0	0	0	1	0	0	0	0	1
12	0	0	0	0	0	0	0	0	0
13	0	1	0	1	0	0	0	0	0
14	1	1	1	1	1	0	1	0	1
15	0	0	0	0	1	0	0	0	0
16	0	0	0	0	0	1	0	1	0
17	1	1	0	1	0	1	0	1	0
18	0	0	0	0	0	0	0	0	0
19	1	1	0	1	1	0	0	0	0
20	0	0	0	0	0	0	0	0	1
21	1	1	0	0	0	1	0	0	0
22	0	0	0	0	0	0	0	0	0
23	1	1	0	0	0	0	0	0	0
24	0	0	0	0	0	0	0	1	0
25	1	1	1	0	0	1	0	1	0
26	1	1	0	0	0	0	1	0	0
27	1	1	0	0	0	0	0	1	0
28	0	0	0	0	0	0	0	0	0
29	1	1	0	1	0	1	0	1	0
30	0	0	0	1	0	0	0	0	0
31	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	1
33	0	0	0	0	0	0	0	0	1
34	0	0	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0
36	1	1	0	0	0	0	0	0	0
37	0	0	0	0	0	0	0	0	0
38	0	0	0	0	0	0	0	0	0
39	0	0	0	0	0	0	0	0	0
40	0	0	0	0	0	0	0	0	0
41	0	0	0	0	0	0	0	0	0
42	0	0	0	0	0	0	0	0	0
43	0	0	0	0	0	0	0	0	0
44	0	0	0	0	0	0	0	0	0
45	0	0	0	0	0	0	0	0	0
46	0	0	0	0	0	0	0	0	0
47	0	0	0	0	0	0	0	0	0
48	0	0	0	0	0	0	0	0	0
49	0	0	0	0	0	0	0	0	0
50	0	0	0	0	0	0	0	0	0
51	0	0	0	0	0	0	0	0	0
52	0	0	0	0	0	0	0	0	0

Table 16 Original medicine prescribed by the doctor.

VI. DISCUSSION OF FIRST FIVE PATIENTS RESULTS WITH DOCTOR:

Original Medicines given by doctor:

A) 1,3,5,6,14.17,19,21,23,25,26,27,29,36
B) 2,3,5,6,14,16,17,21,23,25,26,27,28
C) 1,5,6,14,25
D) 3,5,7,10,11,13,14,17,19,30
E) 3,14,15,19

Medicines given by the Expert system using SVM

A) 1,3,5,6, 16,1718,21,23,25,26,27,28,29
B) 1,3,5,6,16,17,18,21,23,25,26,27,28,29
C) 1,3,5,6,11,14,21,22,23,24,25,26,27
D) 1,3,5,6,13,14,17,2122,23,25,26,27,28
E) 1,2,3,5,14.

Medicines given by Expert System using RBF:

A) 3,5,6,14,16,17,21,23,25,26,27,28,29
B) 1,3,5,6,11,14,16,17,21,22,23,24,25,26,27,28,36
C) 1,3,5,6,14,21,25
D) 1,3,5,6,11,13,14,17,21,25,30
E) 3,5,6,14,15

A. Comparative studies of SVM and RBF methods for medical prescription for heart disease patient

RBF networks and SVM both are examples of non-linear layered feed-forward networks and they are universal approximations. The basic comparison of RBF and SVM NN for the medical prescription for heart disease patient presented in table 14 and table 15.

1) In both the NN model 250 data samples has given as input.
2) The SVM model takes 250 epochs to train it while RBF NN model takes only 225 epochs for the training of the model.
3) If the training performance error is compared SVM NN gives less as compared to RBF NN model.

So Medicines given by the expert system using SVM is not producing the appropriate result as compared with the RBF.

B. So we have taken doctors opinion on the result of RBF as :

Patient A is having major heart attack for which the expert system has provided the medicine no. 1 which is anxiolytic and is given in almost all patients. Medicine no. 16 is beneficial as it reduces the heart rate and thereby reduces workload and improves outcome. Medicine no. 26 can prove useful as it prevents stress erosions/ulcer. The medicine no. 28 is sedative (sleep inducing drug) and is beneficial and if given may help improve outcome.

Patient B Medicine no 1 is alprazolam which is anxiolytic and is given in almost all patients and won't affect the heart patient. Medicine no. 11 should not be prescribed as it is a diuretic and can cause fall in blood pressure/electrolyte imbalance. And is not appropriate and is wrongly given by the expert system. Medicine no 22 is Antioxidant and if given is useful. Medicine no. 24 is antibiotic and is given in presence of infection and dose not affects the cardiac outcome.

Patient C, Medicine 21 is prescribed which has a cardiac remodeling effect and if given improves outcome.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Patient D, Medicine no 1 is alprazolam which is used for its anxiolytic effect is given in almost all patients and wont affect the heart patient. Medicine no. 6 has near to same action as medicine no. 3 which is already prescribed. Expert system has not given medicine no. 7 which is useful in this patient as patients clinical condition has poor heart pumping. Medicine no. 19 is a antipyretic drug (to reduce fever) or as analgesic an if given wont affect the cardiac outcome. Medicine no. 25 is anticlotting medicine which is wrongly given and it should be given in moderate to severe cases after assessing clinical condition of the patient.

Patient E . Medicine no. 5 is not given by the expert system which has cholesterol reducing agent and plays important role for positive outcome. Medicine no. 6 has the same action as medicine no. 3 and is already prescribed.

Medicines given by expert system in few patients are comparatively less and in few patients additional. The system has analyzed 125 sample data and is prescribing 97% accuracy in the medicines as prescribed by the doctor after his clinical assessment. In some cases it is justified but in some cases it depends on the Clinical condition. The additional medicines prescribed may prove beneficial or harmful and vice versa prescribing less medicines (which if essential) can affect the cardiac outcome.

VII. CONCLUSION

In this paper, around 300 patient's information is collected from Sahara Hospital, under supervision of Dr. Abdul Jabbar, (MD Medicine) Sahara Hospital, Roshan Gate, Aurangabad. The collected information is coded, normalized and entered into 13 different excel sub-sheets. All the patients data is trained by using SVM and RBF. Around 50 samples were tested with these two techniques. If the more data set is used for the training the NN model gives more robust results. The analysis model by using SVM and RBF of ANN gives better result for medical prescription for heart disease patient. However, there are several techniques that can improve the speed and performance of the back propagation algorithm, weight initialization, use of momentum and adaptive learning rate. It is found that the result of testing data by using SVM is not satisfactory but the medicines prescribed by the RBF are satisfactory as per the result verified by the doctor. The diagnosis performances of this study shows the advantage of this system. : it is rapid, easy to operate, non-invasive and not expensive. The working prototype model in the field of heart diagnosis can use the system. It also helps for training begineer's doctors and medical students who work in the field of heart diagnosis. In future, this work may be extend using regression technique.

ACKNOWLEDGEMENT

The author thank to Dr. Abdul Jabbar (MD Medicine), Sahara Hospital, Raushan Gate, Aurangabad for supervision of medicine data as well as evaluation of medicine provided by the expert system.

REFERENCES

[1] Fraser, H.S.F., et al., "Differential diagnoses of the heart disease program have better senditivity than Resident Physicians", Tufts-New England Medical Center, Boston, MA(2001).

[2] Long, W., et.al., "Developing a program for Tracking Heart Failure", MIT Lab for Computer Science, Cambridge, MA(2001).

[3] Frase, H.S.F., et.al., "Comparing complex diagnoses a formative evalution of the heart disease program", MIT Lab for Computer Science, Cambridge, MA(2001)

[4] 4. Salem, A.M. Raushdy, M and HodHod, R.A., " A rule based expert system for diagnosis of Heart disease", 8th Internationalconference on soft Computing MENDEL, Brno, Czech Rupublic, June 5-7, PP 258-263(2002).

[5] I. Turkoglu, A. Arslan, E. Ilkay, "An expert system for diagnosis of the heart valve diseases", Expert Systems with Applications vol.23, pp. 229–236, 2002.

[6] I. Turkoglu, A. Arslan, E. Ilkay, "An intelligent system for diagnosis of heart valve diseases with wavelet packet neural networks", Computer in Biology and Medicine vol.33 pp.319–331, 2003.

[7] E. Comak, A. Arslan, I. Turkoglu, "A decision support system based on support vector machines for diagnosis of the heart valve diseases", Computers in Biology and Medicine vol.37, pp.21–27, 2007.

[8] H. Uguz, A. Arslan, I. Turkoglu, "A biomedical system based on hidden Markov model for diagnosis of the heart valve diseases", Pattern Recognition Letters vol.28 pp.395–404, 2007.

[9] A. Sengur, "An expert system based on principal component analysis, artificial immune system and fuzzy k-NN for diagnosis of valvular heart diseases", Computers in Biology and Medicine vol.38 pp.329–338, .2008.

[10] A. Sengur, "An expert system based on linear discriminant analysis and adaptive neuro-fuzzy inference system to diagnosis heart valve diseases", Expert Systems with Applications vol.35 pp.214–222, 2008.

[11] P.I.J. Keeton, F.S. Schlindwein, "Application of Wavelets in Doppler Ultrasound, vol. 17, number 1, MCB University Press, pp. 38–45, 1997.

[12] I.A. Wright, N.A.J. Gough, F. Rakebrandt, M. Wahab, J.P. Woodcock, "Neural network analysis of Doppler ultrasound blood flow signals: a pilot study, Ultrasound in Medicine & Biology vol.23,pp. 683–690, 1997.

[13] B. C. B chan ,F.H.Y. Chan ,F.K. Lam ,P.W.Lui,P.W.F.Poon, ""Fast detection of venous air embolism is Doppler heart sound using the wavelet transform", *IEEE Transaction on Biomedical Engineering* vol. 44 pp.237-245, (1997).

[14] I. Guler, M.K. Kiymik, S. Kara, M.E. Yuksel, "Application of autoregressive analysis to 20 MHz pulsed Doppler data in real time, *International Journal Biomedical Computing* vol.31,pp.247–256, 1992.

[15] http://elearning.najah.edu/OldData/pdfs/Intro%20Expert% 20 Systems%20test-me.co.uk. ppt#262,5, Components of an Expert System.

[16] J. Galindo, P. Tamayo, Credit risk assessment using statistical and machine learning: basic methodology and risk modeling applications, Computational Economics 15 (1 – 2) (2000) 107–143.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[17] D. Michie, D.J. Spiegelhalter, C.C. Taylor, Machine Learning, Neural and Statistical Classification, Elis Horwood, London, 1994.
[18] S.M. Weiss, C.A. Kulikowski, Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Networks, Machine Learning and Expert Systems, Morgan Kaufmann, San Mateo, 1991.
[19] V. Vapnik, The Nature of Statistical Learning Theory, Springer- Verlag, New York, 1995.
[20] M.A. Hearst, S.T. Dumais, E. Osman, J. Platt, B. Scho¨lkopf, Support vector machines, IEEE Intelligent Systems 13 (4) (1998) 18– 28.
[21] M.P. Brown, W.N. Grudy, D. Lin, N. Cristianini, C.W. Sugnet, T.S. Furey, M. Ares, D. Haussler, Knowledge-based analysis of microarray gene expression data by using support vector machines. Proceedings of National Academy of Sciences 97 (1) (2000) 262–267.
[22] T.S. Jaakkola, D. Haussler, Exploiting generative models in discriminative classifiers, in: M.S. Kearns, S.A. Solla, D.A. Cohn (Eds.), Advances in Neural Information Processing Systems, MIT Press, Cambridge, 1998.
[23] A. Zien, G. Ra¨tsch, S. Mika, B. Scho¨lkopf, T. Lengauer, K.-R. Mu¨ller, Engineering support vector machine kernels that recognize translation initiation sites, Bioinformatics 16 (9) (2000) 799– 807.
[24] T. Joachims, Text categorization with support vector machines, Proceedings of the European Conference on Machine Learning (ECML), 1998.
[25] E. Osuna, R. Freund, F. Girosi, Training support vector machines: an application to face detection, Proceedings of Computer Vision and Pattern Recognition, 1997, pp. 130–136.
[26] F.E.H. Tay, L.J. Cao, Modified support vector machines in financial time series forecasting, Neurocomputing 48 (2002) 847– 861.
[27] T. Van Gestel, J.A.K. Suykens, D.-E. Baestaens, A. Lambrechts, G. Lanckriet, B. Vandaele, B. De Moor, J. Vandewalle, Financial time series prediction using least squares support vector machines within the evidence framework, IEEE Transactions on Neural Networks 12 (4) (2001) 809– 821.
[28] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines, Cambridge Univ. Press, Cambridge, New York, 2000.
[29] Broomhead D., & Lowe, D., Multivariable functional interpolation and adaptive networks. Complex Systems, vol.2, pp.321-355, 1988.
[30] Haralambos Sarimveis, Philip Doganis, Alex Alexandridis, "A classification technique based on radial basis function neural networks", Advances in Engineering Software vol.37, pp.218–221, 2006.

AUTHORS PROFILE

Shaikh Abdul Hannan **(Aurangabad, 02-05-1973) is a** Member of IETE, ISTE. He has received the B.Sc. and M.Sc. degree with first class from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad. He has completed M.Phil with first class from Y.C.M.O, Nasik and persuing Ph.D. from North Maharashtra University, Jalgoan. He is lecturer in computer department of Vivekanand College, Aurangabad since last 10 years. He has published more than 40 papers in various international conferences and journals. His has taken Ist prize in International conference at Nagpur (ITBI-2009). He is board member of "Journal of Advanced Research in Computer Engineering: An International Journal" New Delhi, India.

Dr. V. D. Bhagile received the M.C.M degree in 2003 & Ph.D. from *Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, INDIA,* She is recipient of Award of Golden Jubilee University Junior Research fellowship for the years 2007-08 & 2008-09. Presently working as a lecturer & InCharge Principal in Sayali Charitable Trust's College of Computer Science and Information Technology (612), Aurangabad.

Dr. R. R. Manza received M.Sc. degree in 1998 from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra India. He has passed NET/SET in 2002 from Pune university and UGC, New Delhi. He has completed one minor research project on "Development of Biomedical Monitoring System for Diabetes using Nanorobotics in 2008-2009. He has published more and 80 papers in International Journals and conferences. Presently working as a Asst. Professor in Department of Computer Science and I.T., Dr. B.A.M. University, Aurangabad.

Dr. R. J Ramteke received M.Sc. degree in 1995 & Ph.D. from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India. He has taken awards from Teacher Research Fellowship under Faculty Improvement Programme Xth Plan of University Grants Commission, New Delhi during July 2004-July2006 and Finalist for Best Student Paper Award IEEE International Conference on Cybernetics and Intelligent System (CIS-2006), Bangkok, Thailand, on 7th – 9th June 2006. He has published more and 65 papers in International Journals and conferences. Presently working as Reader in Department of Computer Science, North Maharashtra University, Jalgaon.

A Novel Approach for Cardiac Disease Prediction and Classification Using Intelligent Agents

Murugesan Kuttikrishnan
Department of Computer Science and Engineering
Anna University, Chennai, India
murugesan_k3@yahoo.com

Manjula Dhanabalachandran
Department of Computer Science and Engineering
Anna University, Chennai, India
manju@annauniv.edu

***Abstract*— The goal is to develop a novel approach for cardiac disease prediction and diagnosis using intelligent agents. Initially the symptoms are preprocessed using filter and wrapper based agents. The filter removes the missing or irrelevant symptoms. Wrapper is used to extract the data in the data set according to the threshold limits. The classification is based on the prior and posterior probability of the symptoms with the evidence value. Finally the symptoms are classified in to five classes namely absence, starting, mild, moderate, serious. Using the cooperative approach the cardiac problem is solved and verified.**

Keywords- Traditional Chinese Medicine (TCM), Naïve Bayesian Classification (NBC), Bayesian Networks (BN).

I. INTRODUCTION

Intelligent agents are a new paradigm for developing software applications. More than this, agent-based computing has been hailed as 'the next significant breakthrough in software development' (Sargent, 1992), and 'the new revolution in software' (Ovum, 1994). Currently, agents are the focus of intense interest on the part of many sub-fields of computer science and artificial intelligence. Agents are being used in an increasingly wide variety of applications, ranging from comparatively small systems such as email filters to large, open, complex, mission critical systems such as air traffic control. At first sight, it may appear that such extremely different types of system can have little in common. And yet this is not the case: in both, the key abstraction used is that of an agent.

First, an agent is a computer system situated in some environment, and that is capable of autonomous action in this environment in order to meet its design objectives. Autonomy is a difficult concept to pin down precisely, but we mean it simply in the sense • that the system should be able to act without the direct intervention of humans (or other agents), and should have control over its own actions and internal state. It may be helpful to draw an analogy between the notion • of autonomy with respect to agents and encapsulation • with respect to object oriented systems. An object encapsulates some state, and has some control over this state in that it can only be accessed or modified via the methods that the object provides. Agents encapsulate state in just the same way. However, we also think of agents as encapsulating behavior, in addition to state. An object does not encapsulate behavior: it has no control over the execution of methods – if an object x invokes a method m on an object y, then y has no control over whether m is executed or not – it just is. In this sense, object y is not autonomous, as it has no control over its own actions. In contrast, we think of an agent as having exactly this kind of control over what actions it performs. Because of this distinction, we do not think of agents as invoking methods (actions) on agents – rather, we tend to think of them requesting actions to be performed. The decision about whether to act upon the request lies with the recipient. .

An intelligent agent is a computer system that is capable of flexible autonomous action in

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

order to meet its design objectives. By flexible, we mean that the system must be:

- responsive: agents should perceive their environment (which may be the physical world, a user, a collection of agents, the Internet, etc.) and respond in a timely fashion to changes that occur in it.
- proactive: agents should not simply act in response to their environment, they should be able to exhibit opportunistic, goal-directed behavior and take the initiative where appropriate, and Applications of Intelligent Agents
- social: agents should be able to interact, when they deem appropriate, with other artificial agents and humans in order to complete their own problem solving and to help others with their activities.

Hereafter, when we use the term 'agent', it should be understood that we are using it as an abbreviation for 'intelligent agent'. Other researchers emphasize different aspects of agency (including, for example, mobility or adaptability). Naturally, some agents may have additional characteristics, and for certain types of applications, some attributes will be more important than others. However, we believe that it is the presence of all four attributes in a single software entity that provides the power of the agent paradigm and which distinguishes agent systems from related software paradigms – such as object-oriented systems, distributed sysems, and expert systems (see Wooldridge (1997) for a more detailed discussion). By an agent-based system, we mean one in which the key abstraction used is that of an agent. In principle, an agent-based system might be conceptualized in terms of agents, but implemented without any software structures corresponding to agents at all. We can again draw a parallel with object-oriented software, where it is entirely possible to design a system in terms of objects, but to implement it without the use of an object-oriented software environment. But this would at best be unusual, and at worst, counterproductive. A similar situation exists with agent technology; we therefore expect an agent-based system to be both designed and implemented in terms of agents. A number of software tools exist that allow a user to implement software systems as agents, and as societies of cooperating agents. Note that an agent-based system may contain any non-zero number of agents. The multi-agent case – where a system is designed and implemented as several interacting agents, is both more general and significantly more complex than the single-agent case. However, there are a number of situations where the single-agent case is appropriate.

II. RELATED WORK

Traditional diagnosis in TCM requires long experiences and a high level of skill, and is subjective and deficient in quantitative diagnostic criteria. This seriously affects the reliability and repeatability of diagnosis and limits the popularization of TCM. So the focal problem that needs to be solved urgently is to construct methods or models to quantify the diagnosis in TCM.[1] Recently, a few researchers developed some methods and systems to modernize TCM. But most of them are built incorporating totally or partially rulebased reasoning model, which are lack of the feasibility of implementing all possible inference by chaining rules and limits their practical applications in clinical medicines.An attraction tool for managing various forms of uncertainty is Bayesian networks (BNs) [2], [3] which is able to represent knowledge with uncertainty and efficiently performing reasoning tasks.Naive Bayesian classifier (NBC) is a simplified form of BNs that assumes independence of the observations. Some research results [4], [5], [6] have demonstrated that the predictive performance of NBC can be competitive with more complicated classifiers.In this study, a novel computerized diagnostic model based on naive Bayesian classifier (NBC) is proposed. Firstly, a Bayesian network structure is learned from a database of cases [7] to find the symptom set that are dependent on the disease directly.Secondly, the symptom set is utilized as attributes of NBC and the mapping relationships between the symptom set and the disease are constructed.To reduce the dimensionality and improve the prediction accuracy of diagnostic

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

model, symptom selection is requisite.Many feature selection methods, such as filters [8] and wrappers [9], have developed. But the dependency relationships among symptoms and the mapping relationships between symptom and syndrome are not considered in these methods, which are important to diagnosis in TCM.

To lower the influences of irrelative symptoms, the mutual information between each symptom and disease is computed based on information entropy theory [10], which is utilized to assess the significance of symptoms.The paper [11] presents a multiagent system for supporting physicians in performing clinical studies in real time. The multiagent system is specialized in the controlling of patients with respect to their appointment behavior. Novel types of agents are designed to play a special role as representatives for humans in the environment of clinical studies. OnkoNet mobile agents have been used successfully for patient-centric medical problems solving [12].,emerged from a project covering all relevant issues, from empirical process studies in cancer diagnosis/therapy, down to system implementation and validation. In the paper [13], a medical diagnosis multiagent system that is organized according to the principles of swarm intelligence is proposed. It consists of a large number of agents that interact with each other by simple indirect communication.

In the paper [14], a system called Feline composed of five autonomous agents (expert systems with some proprieties of the agents) endowed with medical knowledge is proposed. These agents cooperate to identify the causes of anemia at cats. The paper [39], also presents a development methodology for cooperating expert systems.In the paper [15], a Web-centric extension to a previously developed expert system specialized in the glaucoma diagnosis is proposed. The proposed telehealth solution publishes services of the developed Glaucoma Expert System on the World Wide Web.Each agent member of the CMDS system has problems solving capability and capacity (the notions are defined in [16, 17]). The capacity of an agent Agf (Agf U MDUAS)consists in the amount of problems that can be solved by the agent, using the existent problem solving resources.The cooperative medical diagnosis problems solving by the diagnosis system is partially based on the blackboard-based problem solving [18, 19]. The problem solving by the BMDS system is similar with the situations, when more physicians with different medical specializations plans a treatment to cure an illness that is in an advanced stage. Treatments known to be effective for the curing of the illness in a less advanced stage cannot be applied.

From the entire discussions one can comprehend and classify the medical agent-based IDSS research [20] into two categories, namely Clinical Management and Clinical Research. Clinical Management envelops all clinical systems that are designed to help the doctor with diagnosing and deciding on treatment for medical conditions. Clinical Research on the erstwhile envelopes systems that are used to research facts and connections in attempt to detect new trends and patterns; it covers systems for both diagnosing patients and treating them.

III. PREPROCESSING AND CLASSIFICATION

3.1 Filter agent

Feature selection, as a preprocessing step to machine learning, is effective in reducing dimensionality, removing irrelevant data, in-creasing learning accuracy, and improving result comprehensibility. In this work, we introduce a novel concept, predominant correlation, and propose a faster method which can identify relevant features as well as redundancy among relevant features without pair wise correlation analysis. The efficiency and effectiveness of our method is demonstrated through extensive comparisons with other methods using real-world data of high dimensionality.

3.2 Wrapper agent

In Wrapper based feature selection, the more states that are visited during the search phase of the algorithm the greater the likelihood of finding a

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

feature subset that has a high internal accuracy while generalizing poorly. It removes the irrelevant attributes that are below the threshold value.

3.3 Classifier agent

The classifier agent uses the naïve Bayesian classification algorithm. Bayesian network algorithm is used to classify the collected attributes in to five classes (0-normal,1-starting,2-low,3-mild,4serious). The mutual information between each symptom and disease is computed based on information entropy theory. F and C are symptom and disease. f, c are events of F and C. I(F,C)= P(f, c)logP(f,c)/P(f)P(c) Suppose I0 is the prior entropy of C. I0 = (P(c=1)logp(c=1)+p(c=0)logp(c=0)) The significance of each symptom is calculated by S(F,C)=I(F,C)/I0. All the symptoms are evaluated and ranked by significance index S(F,C).

Input Attributes Documentation

1 id: patient identification number
2 ccf: social security number (I replaced this with a dummy value of 0)
3 age: age in years
4 sex: sex (1 = male; 0 = female)
5 painloc: chest pain location (1 = substernal; 0 = otherwise)
6 painexer (1 = provoked by exertion; 0 = otherwise)
7 relrest (1 = relieved after rest; 0 = otherwise)
8 pncaden (sum of 5, 6, and 7)
9 cp: chest pain type
--Value 1: typical angina
--Value 2: atypical angina
--Value 3: non-anginal pain
--Value 4: asymptomatic
10 trestbps: resting blood pressure (in mm Hg on admission to the hospital)
11 htn
12 chol: serum cholestoral in mg/dl
13 smoke: I believe this is 1 = yes; 0 = no (is or is not a smoker)
14 cigs (cigarettes per day)
15 years (number of years as a smoker)
16 fbs: (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
17 dm (1 = history of diabetes; 0 = no such history)
18 famhist: family history of coronary artery disease (1 = yes; 0 = no)
19 restecg: resting electrocardiographic results
--Value 0: normal
--Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
--Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
20 ekgmo (month of exercise ECG reading)
21 ekgday(day of exercise ECG reading)
22 ekgyr (year of exercise ECG reading)
23 dig (digitalis used furing exercise ECG: 1 = yes; 0 = no)
24 prop (Beta blocker used during exercise ECG: 1 = yes; 0 = no)
25 nitr (nitrates used during exercise ECG: 1 = yes; 0 = no)
26 pro (calcium channel blocker used during exercise ECG: 1 = yes; 0 = no)
27 diuretic (diuretic used used during exercise ECG: 1 = yes; 0 = no)
28 proto: exercise protocol
1 = Bruce
2 = Kottus
3 = McHenry
4 = fast Balke
5 = Balke
6 = Noughton
7 = bike 150 kpa min/min (Not sure if "kpa min/min" is what was written!)
8 = bike 125 kpa min/min
9 = bike 100 kpa min/min
10 = bike 75 kpa min/min
11 = bike 50 kpa min/min
12 = arm ergometer
29 thaldur: duration of exercise test in minutes
30 thaltime: time when ST measure depression was noted
31 met: mets achieved
32 thalach: maximum heart rate achieved
33 thalrest: resting heart rate
34 tpeakbps: peak exercise blood pressure (first of 2 parts)
35 tpeakbpd: peak exercise blood pressure (second of 2 parts)
36 dummy

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

37 trestbpd: resting blood pressure
38 exang: exercise induced angina (1 = yes; 0 = no)
39 xhypo: (1 = yes; 0 = no)
40 oldpeak = ST depression induced by exercise relative to rest
41 slope: the slope of the peak exercise ST segment
--Value 1: upsloping
--Value 2: flat
--Value 3: downsloping
42 rldv5: height at rest
43 rldv5e: height at peak exercise
44 ca: number of major vessels (0-3) colored by flourosopy
45 restckm: irrelevant
46 exerckm: irrelevant
47 restef: rest raidonuclid (sp?) ejection fraction
48 restwm: rest wall (sp?) motion abnormality
0 = none
1 = mild or moderate
2 = moderate or severe
3 = akinesis or dyskmem (sp?)
49 exeref: exercise radinalid (sp?) ejection fraction
50 exerwm: exercise wall (sp?) motion
51 thal: 3 = normal; 6 = fixed defect; 7 = reversable defect
52 thalsev: not used
53 thalpul: not used
54 earlobe: not used
55 cmo: month of cardiac cath (sp?) (perhaps "call")
56 cday: day of cardiac cath (sp?)
57 cyr: year of cardiac cath (sp?)
58 num: diagnosis of heart disease (angiographic disease status)
--Value 0: < 50% diameter narrowing
--Value 1: > 50% diameter narrowing
(in any major vessel: attributes 59 through 68 are vessels)
59 lmt
60 ladprox
61 laddist
62 diag
63 cxmain
64 ramus
65 om1
66 om2
67 rcaprox
68 rcadist
69 lvx1: not used
70 lvx2: not used
71 lvx3: not used
72 lvx4: not used
73 lvf: not used
74 cathef: not used
75 junk: not used
76 name: last name of patient
(I replaced this with the dummy string "name")

IV. EXPERIMENTAL RESULTS

Table: 1 Data Sets Used

S.No.	DATA SET NAME	NO. OF INSTANCES
1	CLEVELAND	303
2	HUNGARIAN	294
3	SWITZERLAND	123
4	LONG BEACH	200

Table: 2 Preprocessed Results-Cleveland Data Set

PID	FILTER	WRAPPER
1	12	13
2	10	12
3	11	7
4	14	10
5	11	13
6	12	12
7	22	15
8	11	12
9	11	10
10	9	10

Table: 3 Preprocessed Results-Hungarian Data Set

PID	FILTER	WRAPPER
1254	10	12
1255	9	11
1256	8	11
1257	7	10
1258	12	13
1259	11	10
1260	13	12
1261	11	10
1262	10	12
1263	13	13

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Table: 4 Preprocessed Results-Swiz Data Set

PID	FILTER	WRAPPER
3001	11	6
3002	10	8
3003	11	9
3004	8	11
3005	9	10
3006	10	11
3007	10	11
3008	12	11
3009	11	11
3010	11	10

Table: 5 Preprocessed Results-Longbeach Data Set

PID	FILTER	WRAPPER
1	11	13
2	10	8
3	11	12
4	13	11
5	14	6
6	12	9
7	11	10
8	12	10
9	6	11
10	11	14

Table: 6 Prior Probability of Symptoms

SYMPTOM	ABSENCE	STARTING	MILD	MODERATE	SERIOUS
PAINLOC	0.3	0.4	0.6	0.7	0.9
PAINEXER	0.1	0.4	0.7	0.85	0.89
RELREST	0.2	0.34	0.4	0.7	0.8
PNCADEN	0.1	0.43	0.5	0.6	0.7
CP	0.3	0.2	0.3	0.4	0.8
TRESTBPS	0.3	0.4	0.45	0.5	0.76
HTN	0.2	0.2	0.4	0.5	0.7
CHOL	0.3	0.3	0.5	0.6	0.778
SMOKE	0.23	0.4	0.45	0.5	0.8
CIGS	0.14	0.4	0.4	0.5	0.9
YEARS	0.15	0.3	0.4	0.466	0.8
FBS	0.16	0.2	0.3	0.4	0.7
DM	0.3	0.3	0.4	0.5	0.888
FAMHIST	0.3	0.333	0.388	0.5	0.677
RESTECG	0.1	0.12	0.2	0.3	0.55
EKGMO	0.2	0.3	0.4	0.5	0.6
EKGDAY	0.3	0.3	0.4	0.488	0.7
EKGYR	0.2	0.3	0.395	0.4	0.9
DIG	0.23	O.3	0.35	0.455	0.8

TABLE: 7 CLASSIFICATION RESULTS

S.NO.	DATA SET	ABSENCE	STARTING	MILD	MODERATE	SERIOUS
1	CLEVELAND	164	55	36	35	13
2	HUNGARIAN	188	37	26	28	15
3	SWITZERLAND	8	48	32	30	5
4	LONG BEACH	51	56	41	42	10

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

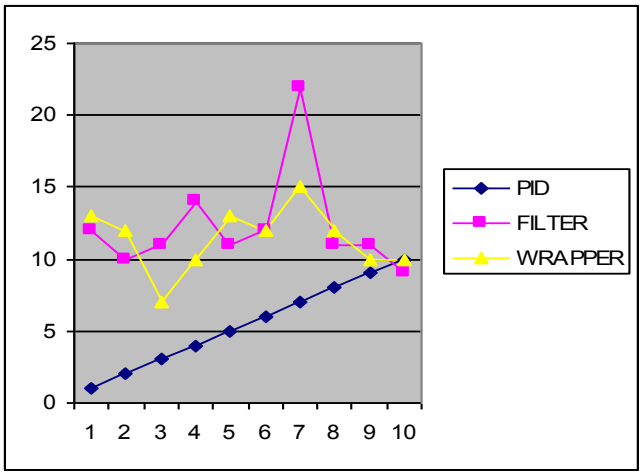


Figure: 1 Preprocessed Chart for Cleveland Data Set

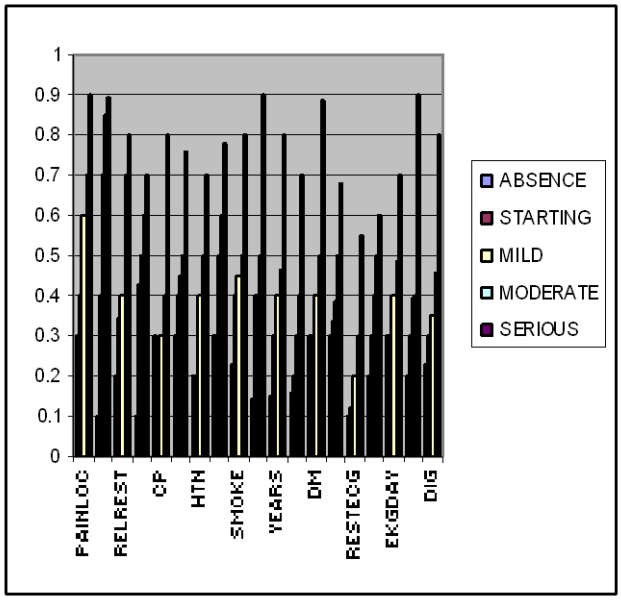


Figure: 2 Classification Chart for all Data Sets

V. CONCLUSIONS AND FUTURE WORK

In the current work, input symptom values are preprocessed using filter and wrapper approach. Retained symptom values are used to classify cardiac patients in to five classes viz. Normal, Starting, Mild, Moderate and Serious. Classified values determines severity of the cardiac disease. Future work represents Heuristic seed selection to increase classification accuracy over the supervised naïve Bayesian classification. In this approach five seeds are selected for five classes. Future work represents diagnosis of cardiac patients using the co-operative elaboration algorithm. It uses contract net protocol that allows autonomy and sharing among the agents to solve all cardiac problems. Diagnosis results will be accurate in all conditions and then intelligent based decision support system is used to retrieve the patient information from the database.

VI. ACKNOWLEDGEMENT

Primarily, my gratitude goes to Dr. D.Manjula, A.P, DCSE, Anna University Chennai, who guided me to carry out this research with interest and involvement. With profound reverence and high regards, I thank Dr. P. Narayanasamy, Professor and HOD, DISE, Anna University Chennai and Dr.R.Krishnamoorthy, Professor and Dean, Anna University Trichy, for the motivation provided by them during this research work which enabled me to complete this work successfully.

REFERENCES

[1] B. Ma, "Expert Systems and Knowledge Bases in Traditional Chinese Medicine", Beijing Press, Beijing, 1994
.

[2] J. Pearl, "Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference" Morgan Kaufmann Publishers, Inc. San Francisco, California, 1988.

[3] J. Cheng,, C. Hatzis,, H. Hayashi, M.A. Krogel, S. Morishita, D. Page, J. Sese, "KDD Cup 2001 report" SIGKDD Explor. Newsl, 2002. (2), pp.47-64.

[4] W. Buntine, "Learning classification rules using Bayes", in Proc. 6th Int. Workshop Machine Learning, 1989, pp. 94–96.

[5] B. Cestnik, "Estimating probabilities: A crucial task in machine learning," in Proc. Euro. Conf. Artif. Intell., 1990, pp. 147–149.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[6] P. Langley, W. Iba, and K. Thompson, "An analysis of 1992", pp. 223–228.

[7] J. Cheng, R. Greiner, J. Kelly, D. A. Bell, W. Liu, "Learning Bayesian Networks from Data: An Information-Theory Based Approach", Artificial Intelligence, 2002, Vol.137, pp.43-90.

[8] G. Isabelle, "An Introduction to Variable and Feature Selection", Journal of Machine Learning Research, Vol. 3, pp.1157-1182.

[9] R. Kohavi, G. John, "Wrappers for Feature Selection",Artificial Intelligence, Vol. 97, No. 1, 1997, pp.273-324.

[10] C. E. Shannon, W. Weaver, "The Mathematical Theory of Communication.Urbana, IL: University of Illinois Press,1949.

[11] H. Myritz, G. Lindemann, G. Zahlmann, Hans-Dieter Burkhard. "Patient scheduling in clinical studies with multiagent techniques". Proceedings of the 2nd International Workshop on Multi-Agent Systems for Medicine and Computational Biology, Hakodate, Japan, pages 87–103, 2006

[12] S. Kirn. Ubiquitous healthcare: "The onkonet mobile agents architecture". Proceedings of the 3.rd International Conference Netobjectdays. Objects, Components, Architectures,Services, and Applications for a NetworkedWorld, Aksit, M.,Mezini, M., Unland, R. (Eds.), Springer-Verlag, Germany, LNCS, (2591), 2003.

[13] R. Ulieru, M. Unland. "A stigmergic approach to medical diagnosis".Proceedings of the 2nd International Workshop on Multi-Agent Systems for Medicine and Computational Biology, Hakodate, Japan, pages 87–103, 2006.

[14] G. Weiss. (Ed.)" Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence" MIT Press Cambridge, Massachusetts London, 2000.

[15] M. Wooldridge, G. M. P. O'Hare, R. Elks. Feline – "a case study in the design and implementation of a co-operating expert system". Proceedings of the International Conference on Expert Systems and their Applications, France, 1991.

[16] A. Ulieru, M. Grabelkovsky. "Telehealth approach for glaucoma progression monitoring."International Journal: Information Theories and Applications, (10):326–329, 2005.

[17] B. Iantovics, C. Chira, D. Dumitrescu. "Principles of the Intelligent Agents"Casa Cartii de Stiinta Press, Cluj-Napoca, 2007.

[18] J. Ferber. "Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence" Addison Wesley, 1999.

[19] V. Jagannathan, R. Dodhiawala, L. S. Baum. (Eds.) "Blackboard Architectures and Application". Academic Press, San Diego, 1989.

[20] Murugesan.K., Manjula. D., "Applications of Intelligent Agents in Health Care Systems using Decision Support Systems," Proceedings of International Conference on Digital Factory (ICDF 2008), CIT, Coimbatore, India, Aug.1113, 2008: p.173.

AUTHORS PROFILE

K. Murugesan**,** is pursuing his Ph.D in Computer Science from Anna University and also working for Anna University as a Superintendent. His research interests include Data Mining, Multi agent based systems etc.,

D.Manjula is working for Anna University as an Assistant Professor. Her research interests include Natural Language Processing, Text Mining, Artificial Intelligence, Databases and Parallel Computing.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Document Image Segmentation Based On Gray Level Co- Occurrence Matrices and Feed Forward Neural Network

S.Audithan
Research Scholar, Dept of CSE,
Annamalai University,
Annamalai NagarTamil Nadu, India
sarabar36@rediffmail.com

Dr. RM. Chandrasekaran
Professor, Dept of CSE
Annamalai University,
Annamalai Nagar Tamil Nadu, India
aurmc@sify.com

***Abstract*—This paper presents a new method for extracting text region from the document images employing the combination of gray level co-occurrence matrices (GLCM) and artificial neural networks (ANN). We used GLCM features quantitatively to evaluate textural parameters and representations and to determine which parameter values and representations are best for extracting text region. The detection of text region is achieved by extracting the statistical features from the GLCM of the document image and these features are used as an input of neural network for classification. Experimental results show that our method gives better text extraction than other methods.**

Keywords-component;Document segmentation, GLCM, ANN, Haralick features

I. INTRODUCTION

The extraction of textual information from document images provides many useful applications in document analysis and understanding, such as optical character recognition, document retrieval, and compression. To-date, many effective techniques have been developed for extracting characters from monochromatic document images. The document image segmentation is an important component in the document image understanding.

An efficient and computationally fast method for segmenting text and graphics part of document images based on textural cues is presented in [1]. The segmentation method uses the notion of multi scale wavelet analysis and statistical pattern recognition. M band wavelets are used which decompose an image into M x M band pass channels. Information from the table of contents (TOC) pages can be extracted to use in

document database for effective retrieval of the required pages. Fully automatic identification and segmentation of table of contents (TOC) page from scanned document is discussed in [2].

Character segmentation is the first step of OCR system that seeks to decompose a document image into a sequence of sub images of individual character symbols. Segmentation of monochromatic document images into four classes are presented in [3]. They are background, photograph, text, and graph. Features used for classification are based on the distribution patterns of wavelet coefficients in high frequency bands.

Probabilistic latent semantic analysis (pLSA) model is presented in [4]. The pLSA model is originally developed for topic discovery in text analysis using "bag-of-words" document representation. The model is useful for image analysis by "bag-of-visual words" image representation. The performance of the method depends on the visual vocabulary generated by feature extraction from the document image. Kernel-based methods have demonstrated excellent performances in a variety of pattern recognition problems. The kernel-based methods and Gabor wavelet to the segmentation of document image is presented in [5]. The feature images are derived from Gabor filtered images. Taking the computational complexity into account, the sampled feature image is subjected to Spectral Clustering Algorithm (SCA). The clustering results serve as training samples to train a Support Vector Machine (SVM).

The steerable pyramid transform is presented in [6]. The features extracted from pyramid sub bands serve to locate and classify regions into text and non text in some noise infected, deformed, multilingual, multi script document images. These documents contain tabular structures, logos, stamps, handwritten text blocks, photos etc. A novel scheme for the extraction of textual areas of an image using globally matched wavelet filter is presented in [7]. A clustering based technique has been devised for estimating globally matched wavelet filters using a collection of ground truth images. This work extended to text extraction for the segmentation of document images into text, background, and picture components.

A classical approach in the segmentation of Canonical Syllables of Telugu document images is proposed in [8]. The model consists of zone separation and component extraction phases as independent parts. The relation between zones and components is established in the segmentation process of canonical syllable. The segmentation efficiency of the proposed model is evaluated with respect to the canonical groups.

It [9] presents a new method for extracting characters from various real-life complex document images. It applies a

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

multi-plane segmentation technique to separate homogeneous objects including text blocks, non-text graphical objects, and background textures into individual object planes. It consists of two stages - automatic localized multilevel thresholding, and multi-plane region matching and assembling. Two novel approaches for document image segmentation are presented in [10]. In text line segmentation a Viterbi algorithm is proposed while an SVM-based metric is adopted to locate words in each text line.

Gray-level co-occurrence matrices (GLCM) to quantitatively evaluate textural parameters and representations and to determine which parameter values and representations are best for mapping sea ice texture. In addition, it [11] presents the three GLCM implementations and evaluated them by a supervised Bayesian classifier on sea ice textural contexts. Texture is one of the important characteristics used in identifying objects or region of interest in an image, whether the image to be a photomicrograph, an aerial photograph, or a satellite image. Some easily computable textural features are presented in [12].

II. METHODOLOGY

A. GLCM

Gray level co occurrence matrix (GLCM) is the basis for the Haralick texture features [12]. This matrix is square with dimension N_g, where N_g is the number of gray levels in the image. Element [*i*,*j*] of the matrix is generated by counting the number of times a pixel with value *i* is adjacent to a pixel with value *j* and then dividing the entire matrix by the total number of such comparisons made. Each entry is therefore considered to be the probability that a pixel with value *i* will be found adjacent to a pixel of value *j*.

$$G = \begin{bmatrix} p(1,1) & p(1,2) & \dots & p(1,N_g) \\ p(2,1) & p(2,2) & \dots & p(2,N_g) \\ . & . & \dots & . \\ . & . & \dots & . \\ p(N_g,1) & p(N_g,2) & \dots & p(N_g,N_g) \end{bmatrix}$$

Since adjacency can be defined to occur in each of four directions in a 2D, square pixel image (horizontal, vertical, left and right diagonals as shown in Figure 1, four such matrices can be calculated.

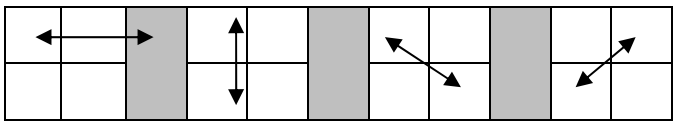


Figure: 1 Four directions of adjacency as defined for calculation of the Haralick texture features.

The Haralick statistics are calculated for co-occurrence matrices generated using each of these directions of adjacency. In our proposed system, based on the gray level occurrence matrix, 10 features are selected for text extraction.

B. ANN

An artificial neuron is a computational model inspired in the natural neurons. Natural neurons receive signals through synapses located on the dendrites or membrane of the neuron. When the signals received are strong enough (surpass a certain threshold), the neuron is activated and emits a signal though the axon. This signal might be sent to another synapse, and might activate other neurons. The complexity of real neurons is highly abstracted when modeling artificial neurons. These basically consist of inputs (like synapses), which are multiplied by weights (strength of the respective signals), and then computed by a mathematical function which determines the activation of the neuron. Another function (which may be the identity) computes the output of the artificial neuron (sometimes in dependence of a certain threshold). ANNs combine artificial neurons in order to process information.

A single layer network has severe restrictions the class of tasks that can be accomplished is very limited. The limitation is overcome by the two layer feed forward network. The central idea behind this solution is that the errors for the units of the hidden layer are determined by back propagating the errors of the units of the output layer. For this reason the method is often called the back propagation learning rule. Back propagation can also be considered as a generalization of the delta rule for nonlinear activation functions and multi layer networks.

A feed forward network has a layered structure. Each layer consists of units which receive their input from units from a layer directly below and send their output to units in a layer directly above the unit. There are no connections within a layer. The N_i inputs are fed into the first layer of $N_{h,i}$ hidden units. The input units are merely “fan out” units no processing takes place in these units. The activation of a hidden unit is a function F_i of the weighted inputs plus a bias. The output of the hidden units is distributed over the next layer of $N_{h,2}$ hidden units until the last layer of hidden units of which the outputs are fed into a layer of N_o output units as shown in Figure 2. In our proposed method a one input layer, two hidden layers and one output layer feed forward neural network is used.

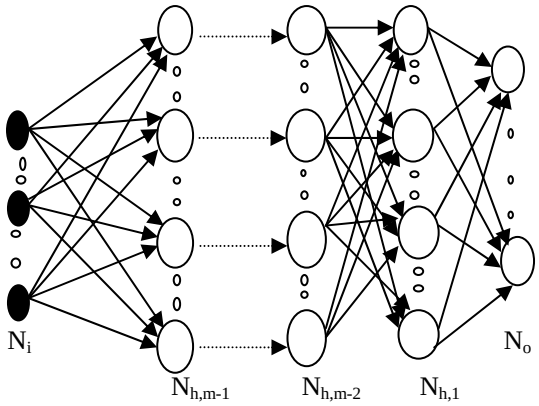


Figure 2. A multi layer networks with m layers of input

III. PROPOSED SYSTEM

The block diagram of the proposed text extraction from document images is shown in Figure 3, where Figure

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

3(a) depicts the Feature Extraction Phase while classification Phase is shown in the Figure 3(b).

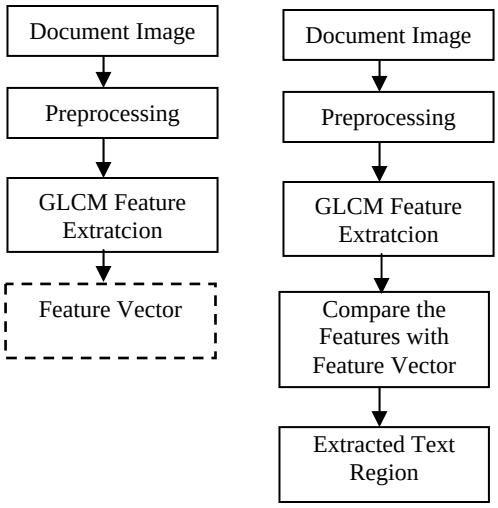


Figure 3: Our Proposed Text extraction method
(a) Feature Extraction Phase (b) Classification Phase

A. *Preprocessing*

Image pre-processing is the term for operations on images at the lowest level of abstraction. The aim of pre-processing is an improvement of the image data that suppresses undesired distortions or enhances some image features relevant for further processing and analysis task. First, the given document image is converted into gray scale image. Then Adaptive Histogram Equalization (AHE) is applied to enhance the contrast of the image. AHE computes the histogram of a local window centered at a given pixel to determine the mapping for that pixel, which provides a local contrast enhancement.

B. *Feature Extraction*

Feature extraction is an essential pre-processing step to pattern recognition and machine learning problems. It is often decomposed into feature construction and feature selection. In our approach, GLCM features are used as a feature vector to extract the text region from the document images. In the following section gives the overview of feature extraction of the text region and the graphics /image part.

The input to the feature extraction module is the document images having text and graphics/image part. A 20X20 non overlapping window is used to extract the features. The GLCM features are extracted from the text region and the graphics parts and stored separately for training phase in the classifier stage.

The following 10 GLCM features are selected for the feature extraction phase. Let $P(i,j)$ be the (i,j) the entry in a normalized GLCM. The mean and standard deviations for thr rows and columns of the matrix are

$$\mu_x = \sum_i \sum_j i.P(i,j) \qquad (1)$$

$$\mu_y = \sum_i \sum_j j.P(i,j) \qquad (2)$$

N is the number of gray levels in the image.

1) Contrast

To emphasize a large amount of contrast, create weights so that the calculation results in a larger figure when there is great contrast.

$$f1 = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i,j)^2 P(i,j) \qquad (3)$$

When i and j are equal, the cell is on the diagonal and (i-j) =0. These values represent pixels entirely similar to their neighbor, so they are given a weight of 0. If i and j differ by 1, there is a small contrast, and the weight is 1. If i and j differ by 2, contrast is increasing and the weight is 4. The weights continue to increase exponentially as (i-j) increases.

2) Cluster prominence

Cluster Prominence, represents the peakedness or flatness of the graph of the co-occurrence matrix with respect to values near the mean value.

$$f2 = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i+j-\mu_x-\mu_y)^4 P(i,j) \qquad (4)$$

3) Cluster Shade

Cluster Shade represents the lack of symmetry in an image and is defined by (5).

$$f3 = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i+j-\mu_x-\mu_y)^3 P(i,j) \qquad (5)$$

4) Dissimilarity

In the Contrast measure, weights increase exponentially (0, 1, 4, 9, etc.) as one moves away from the diagonal. However in the dissimilarity measure weights increase linearly (0, 1, 2,3 etc.).

$$f4 = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |i,j| P(i,j) \qquad (6)$$

5) Energy

Angular second moment (ASM) and Energy uses each P_{ij} as a weight for itself. High values of ASM or Energy occur when the window is very orderly.

$$ASM = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P(i,j)^2 \qquad (7)$$

The square root of the ASM is sometimes used as a texture measure, and is called Energy.

$$f5 = \sqrt{ASM} \qquad (8)$$

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

6) Entropy

Entropy is a notoriously difficult term to understand; the concept comes from thermodynamics. It refers to the quantity of energy that is permanently lost to heat ("chaos") every time a reaction or a physical transformation occurs. Entropy cannot be recovered to do useful work. Because of this, the term is used in non technical speech to mean irremediable chaos or disorder. Also, as with ASM, the equation used to calculate physical entropy is very similar to the one used for the texture measure.

$$f6 = -\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P(i,j)\log_2 P(i,j) \quad (9)$$

1) Maximum Propability

$$f7 = MAX\big(P(i,j)\big) \quad (10)$$

2) Sum Entropy

$$f8 = -\sum_{i=0}^{2(N-1)} P_{i+j}(i)\log\big\{P_{i+j}(i)\big\} \quad (11)$$

3) Difference variance

$$f9 = \sum_{i=0}^{N-1} i^2 P_{i-j}(i) \quad (12)$$

4) Difference entropy

$$f10 = -\sum_{i=0}^{N-1} P_{i-j}(i)\log\big\{P_{i-j}(i)\big\} \quad (13)$$

In the classification stage, the extracted GLCM features are given to the two layers feed forward neural network to classify text and non text region.

IV. Experimental Results

The performance of the proposed system is evaluated in this section. A set of 50 document images are employed for experiments on performance evaluation of text extraction. These test images are scanned from the newspapers and magazines. They are transformed into gray scale images and preprocessed by adaptive histogram equalization for feature extraction. Document image and the text extracted from the given input image are shown in Figure 4(a), 5(a) and 4(b),5(b) respectively.

Figure 4: (a) Input image

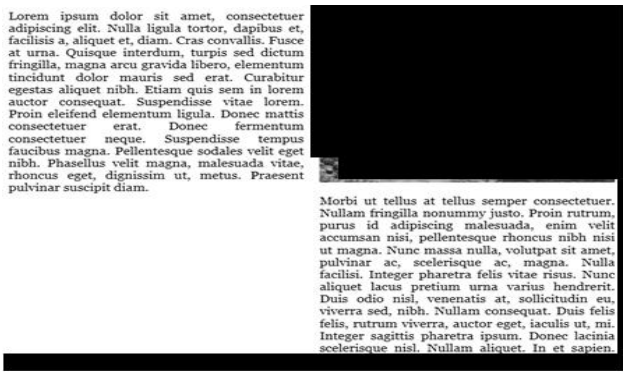


Figure 4: (b) Segmented Result

Figure 5: (a) Input Image

Figure 5: (b) Segmented Result

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

In table 1, we present the feature vector for text samples of 20x20 windows and Figure 6. Shows the Pictorial representations of feature set for sample text regions. In table 2, we present the feature vector for non text samples of 20x20 windows and Figure 7 Shows the Pictorial representations of feature set for sample non text regions.

TABLE I. FEATURE SET FOR TEXT REGION SAMPLES

Feature set	Sample Text 1	Sample Text 2	Sample Text 3	Sample Text 4
				
F1	4.6819	5.9833	4.9583	5.6819
F2	227.69	254.88	222.69	191.23
F3	-25.84	-24.57	-23.46	-18.05
F4	1.2986	1.5277	1.3333	1.5652
F5	0.2308	0.1573	0.2210	0.1274
F6	2.5679	2.8408	2.6287	2.9532
F7	0.4694	0.3777	0.4597	0.3319
F8	1.8594	2.0552	1.8912	2.0778
F9	1.8594	5.9833	4.9583	5.6819
F10	1.4964	1.5856	1.4843	1.6306

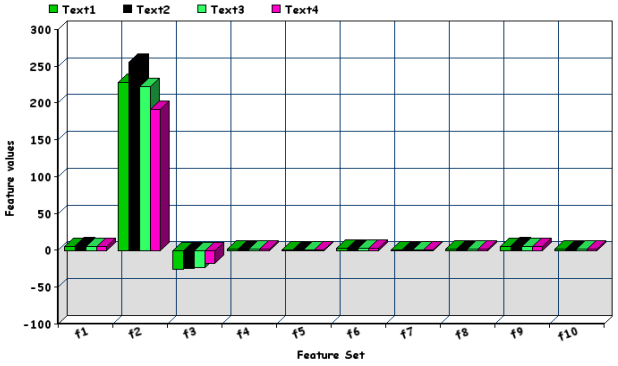


Figure 6: Pictorial representations of feature set for Sample text regions

TABLE II. FEATURE SET FOR NON TEXT REGION SAMPLES

Feature set	Sample non Text 1	Sample non Text 2	Sample non Text 3	Sample non Text 4
				
F1	0.8333	3.9083	1.1416	2.7263
F2	35.725	64.613	27.882	88.857
F3	-6.3919	1.1596	3.0661	-2.0649
F4	0.4305	1.5611	0.6833	1.2152
F5	0.4239	0.0335	0.1259	0.0560
F6	1.5641	3.5657	2.3313	3.2723
F7	0.6361	0.0611	0.2236	0.1166
F8	1.2107	2.2177	1.6982	2.1497
F9	0.8333	3.9083	1.1416	2.7263
F10	0.8674	1.5356	1.0457	1.3863

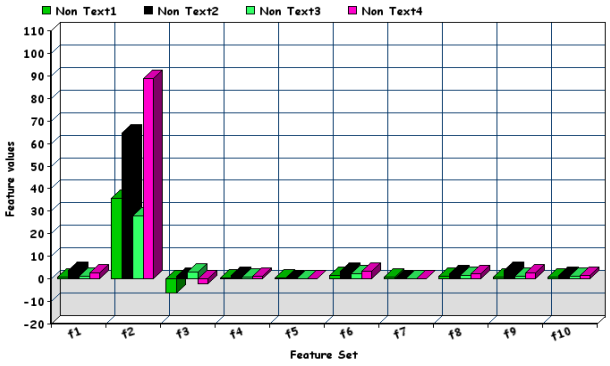


Figure 7: Pictorial representations of feature set for sample Non text regions

V. CONCLUSION

In this paper, we have presented a novel technique for extracting the text part based on feed forward neural network by using GLCM. The GLCM features are serving as training samples to train the neural network. The performance depends on the representation generated by feature extraction from the images. The experimental results indicate the proposed algorithm can effectively process a variety of document images with fonts, structures and components. When the method is used to select engines for document image recognition, it is preferable that the computation time is much faster with minimal error rate.

REFERENCES

[1] Mausumi Acharyya and Malay K.Kundu, “Document Image Segmentation UsingWavelet Scale–Space Features”, IEEE transactions on circuits and systems for video technology,Vol.12, No.12, December2002J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] S.Mandal, S.P.Chowdhury, A.K.Das and Bhabatosh Chanda, “Automated Detection and Segmentation of Table of Contents Page from Document Images”, IEEE Seventh International Conference on Document Analysis and Recognition on 2003.

[3] JiaLi and Robert M.Gray, “Context Based Multiscale Classification of Document Images Using Wavelet Coefficient Distributions”, IEEE transactions on image processing, Vol.9, No.9, September2000.R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[4] Takuma Yamaguchi, and Minoru Maruyama, “Feature extraction for document image segmentation by pLSA model”, The Eighth IAPR Workshop on Document Analysis Systems.2008

[5] Yu-Long Qiao, Zhe-Ming Lu, Chun-Yan Song and Sheng-He Sun, “Document Image Segmentation Using Gabor Wavelet and Kernel-based Methods”, Systems and Control in Aerospace and Astronautics, 2006.

[6] Mohamed Benjelil, Slim Kanoun, Rémy Mullot and Adel M. Alimi, “Steerable pyramid based complex documents images segmentation”, IEEE 10th International Conference on Document Analysis and Recognition on 2009.

[7] SunilKumar, RajatGupta, NitinKhanna and SantanuChaudhury, “Text Extraction and Document Image Segmentation Using Matched Wavelets and MRF Model”, IEEE transactions on image processing, vol.16,no.8, August2007.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[8] L. Pratap Reddy, A.S.C.S. Sastry and A.V. Srinivasa Rao, “Canonical Syllable Segmentation of Telugu Document Images”, TENCON 2008-IEEE Region 10 Conference .

[9] Yen-Lin Chen and Bing-Fei Wu, “Text Extraction from Complex Document Images Using the Multi-plane Segmentation Technique”, IEEE International Conference on Systems, Man, and Cybernetics October 8-11, 2006.

[10] Themos Stafylakis, Vassilis Papavassiliou, Vassilis Katsouros and George Carayannis, “Robust text-line and word segmentation for handwritten documents images”, IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008

[11]]. Leen-Kiat Soh and Costas Tsatsoulis, “Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices”, IEEE transactions on geosciences and remote sensing, vol.37, no.2, march1999.

[12] Robert M.Haralick, K.Shanmugam and Hak Dinstein, “ Textural Features for image Classification”, IEEE transcation on system, man and cybernetics vol no 6, November 1973.

[13] Robert M. Haralick,``Statistical and structural approaches to texture,'' *Proc. IEEE*, vol. 67, no. 5, pp. 786-804, 1979.

AUTHORS PROFILE

S.Audithan is currently a reseaech scholar at the Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar, Tamilnadu, India.. He received his BE degree from BharathiDhasan University in 2000 and ME degree from Annamalai University in 2006. He worked as a network engineer in RBCOMTEC at Hydrabad from 2000 to 2003. Ha has presented and published more than 5 papers in conferences and journals. His research interests include Image processing,Network Security, and Artificial Intelligence.

Dr.RM.Chandrasekaran is currently working as a Professor at the Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar, Tamilnadu, India. From 1999 to 2001 he worked as a software consultant in Etiam, Inc, California, USA. He received his Ph.D degree in 2006 from Annamalai University, Chidambaram. He has conducted workshops and conferences in the area of Multimedia, Business Intelligence, Analysis of Algorithms and Data Mining. Ha has presented and published more than 32 papers in conferences and journals and is the co-author of the book Numerical Methods with C++ Program(PHI,2005). His research interests include Data Mining, Algorithms,Image processing and Mobile Computing. He is life member of the Computer Society of India, Indian Society for Technical Education, Institute of Engineers and Indian Science Congress Assciation.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

IJCSIS REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA

Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia

Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA

Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway

Assoc. Prof. N. Jaisankar, VIT University, Vellore,Tamilnadu, India

Dr. Amogh Kavimandan, The Mathworks Inc., USA

Dr. Ramasamy Mariappan, Vinayaka Missions University, India

Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China

Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA

Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico

Dr. Neeraj Kumar, SMVD University, Katra (J&K), India

Dr Genge Bela, "Petru Maior" University of Targu Mures, Romania

Dr. Junjie Peng, Shanghai University, P. R. China

Dr. Ilhem LENGLIZ, HANA Group - CRISTAL Laboratory, Tunisia

Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India

Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain

Prof. Dr.C.Suresh Gnana Dhas, Anna University, India

Mrs Li Fang, Nanyang Technological University, Singapore

Prof. Pijush Biswas, RCC Institute of Information Technology, India

Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia

Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India

Mr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand

Mr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India

Mr. Hayder N. Jasem, University Putra Malaysia, Malaysia

Mr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India

Mr. R. S. Karthik, C. M. S. College of Science and Commerce, India

Mr. P. Vasant, University Technology Petronas, Malaysia

Mr. Wong Kok Seng, Soongsil University, Seoul, South Korea

Mr. Praveen Ranjan Srivastava, BITS PILANI, India

Mr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong

Mr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia

Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan

Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria

Dr. Riktesh Srivastava, Skyline University, UAE

Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia

Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt
and Department of Computer science, Taif University, Saudi Arabia

CALL FOR PAPERS
International Journal of Computer Science and Information Security
IJCSIS 2010
ISSN: 1947-5500
http://sites.google.com/site/ijcsis/

International Journal Computer Science and Information Security, now at its sixth edition, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2010 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

Track A: Security

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc,), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on
its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

Track B: Computer Science

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embeded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail ijcsiseditor@gmail.com. Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at http://sites.google.com/site/ijcsis/authors-notes .

ISSN 1947 5500